



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

Identificación de marcadores moleculares, SNPs,
asociados a características morfológicas de *Bixa
orellana* L.

Tesis que presenta

Ana Lucía López Gurgua

En opción al título de

MAESTRO EN CIENCIAS

(Ciencias Biológicas: Biología Integrativa)

Mérida, Yucatán, México

2024

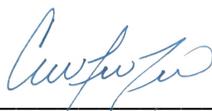
CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS



RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis Ana Lucía López Gurgua titulado **Identificación de marcadores moleculares, SNPs, asociados a características morfológicas de *Bixa orellana* L.**, fue realizado en la Unidad de Biología Integrativa, en la línea de investigación de Metabolismo Secundario e Ingeniería Metabólica, en el laboratorio 27 del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección de la Dra. Renata Lourdes Barbara Rivera Madrid y Codirección del Dr. Jorge Humberto Ramírez Prado, dentro de la opción de Bioquímica y Biología Molecular, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente



Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 26 de junio de 2024

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en las secciones de: Materiales y Métodos, Resultados y Discusión de este documento, proviene de las actividades de investigación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que, a razón de lo anterior, y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley Federal de Protección a la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y se regirán, en todo caso, por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley Federal de Protección a la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.



Ana Lucía López Gurgua

Este trabajo se llevó a cabo en la Unidad de Bioquímica y Biología Molecular del Centro de Investigación Científica de Yucatán, A.C., y forma parte del proyecto titulado “Determinación de la función de los tres genes de la ruta de biosíntesis de bixina en organismos heterólogos” en el proyecto fronteras de la ciencia No.1716 bajo la dirección de la Dra. Renata Rivera Madrid.

AGRADECIMIENTOS

Al CONAHCYT por la beca otorgada (No CVU/Becario: 1197493), la cual me permitió realizar y concluir con éxito mis estudios de maestría.

Al CICY por la instalaciones y facilidades al conocimiento del cuál gozo ahora.

A la Dra. Renata Rivera Madrid por su apoyo, tiempo y motivación al dirigir esta tesis.

Al Dr. Jorge Humberto Ramírez Prado por su tiempo y los conocimientos compartidos.

A mi comité tutorial y revisor de tesis conformado por los Drs. Ruby Valdez Ojeda, June Kilpatric Sipson Willianson y Hugo Antonio Tovar Romero por su tiempo, comentarios y preguntas para poder mejorar esta tesis.

A la M. C. Margarita Aguilar Espinosa por su tiempo y dedicación en enseñarme los experimentos realizados en el laboratorio.

A todos mis compañeros que conforman el grupo de la Dra Renata, Gaby Rivero, Lupita Hernández, Rocío Tamayo, Diego Torres, Rodolfo Pech y, Neimi, por sus comentarios acerca del trabajo y el apoyo moral.

A mis amigos Gaby Vázquez, Diana Sierra y Stefano Decler por la motivación, apoyo y por hacer del CICY un lugar que siempre quiero recordar.

DEDICATORIAS

A mis padres Miguel A. López Anaya y Lucía G. Gurgua de los Santos, quienes me dieron el privilegio de estudiar lo que me apasiona, por confiar en mi y siempre instruirme en amor y sabiduría.

A mi hermana Marcela López Gurgua, mi mejor amiga, quién en días difíciles una palabra de ella me levantaba.

A mi amado Mario Barrientos, quien me motivó, aconsejó, abrazó y festejó durante este importante paso en mi vida.

A Dios porque al final es solo gracias a Él, fue por medio de su creación que descubrí mi pasión por la ciencia.

“Pues la tierra se llenará
del conocimiento de la gloria del Señor
como las aguas cubren el mar.”
Habacuc 2:14 (LBLA)

INDICE

INTRODUCCIÓN.....	1
CAPÍTULO I	3
ANTECEDENTES	3
1.1.GENERALIDADES DE <i>BIXA ORELLANA</i>	3
1.1.1.DESCRIPCIÓN BOTÁNICA	3
1.1.2.IMPORTANCIA SOCIOECONÓMICA	6
1.2.CAROTENOIDES	6
1.2.1. BIOSÍNTESIS DE CAROTENOIDES EN PLANTAS SUPERIORES	7
1.2.1.1. VÍAS PRECURSORAS	7
1.2.1.2. BIOSÍNTESIS DE CAROTENOIDES	9
1.3. APOCAROTENOIDES.....	10
1.4. BIXINA	12
1.4.1. BIOSÍNTESIS DE BIXINA.....	12
1.5. MORFOLOGÍA Y GENÉTICA DEL FRUTO	14
1.6. MARCADORES MOLECULARES	15
1.6.1. POLIMORFISMO DE UN SOLO NUCLEÓTIDO (SNP).....	15
1.6.2. ANTECEDENTES DEL USO DE MARCADORES MOLECULARES	16
JUSTIFICACIÓN	18
HIPÓTESIS.	19

OBJETIVO GENERAL	20
OBJETIVOS ESPECÍFICOS	20
ESTRATEGIA EXPERIMENTAL	21
CAPÍTULO II	22
MATERIALES Y MÉTODOS	22
2.1. ANÁLISIS BIOINFORMÁTICO	22
2.1.1. INFORMACIÓN DE LOS TRANSCRIPTOMAS	22
2.1.2. ANÁLISIS DE CALIDAD Y LIMPIEZA DE LECTURAS	22
2.1.3. MAPEO DE LECTURAS	23
2.1.4. BÚSQUEDA DE VARIANTES MEDIANTE EL MÉTODO VARIANT CALLING (LLAMADO DE VARIANTES)	23
2.1.5. ANÁLISIS DE POLIMORFISMOS DE UN SOLO NUCLEÓTIDO	23
2.2. ESTABLECIMIENTO DE LA TÉCNICA TETRA PRIMER ARMS-PCR	23
2.2.1. MATERIAL VEGETAL	23
2.2.2. EXTRACCIÓN DE DNA	25
2.2.3. TETRA PRIMER ARMS-PCR	25
CAPÍTULO III	27
RESULTADOS	27
3.1. ANÁLISIS BIOINFORMÁTICO	27
3.1.1 ANÁLISIS DE CALIDAD Y MAPEO DE LECTURAS DE TRANSCRIPTOMAS DE TRES ACCESIONES DE <i>B. ORELLANA</i>	27
3.2 LLAMADO DE VARIANTES DE TRANSCRIPTOMAS DE TRES ACCESIONES DE <i>B. ORELLANA</i> .	29

3.3 SELECCIÓN ALEATORIA Y RELACIÓN GENOTIO/FENOTIPO DE POLIMORFISMOS DE UN SOLO NUCLEÓTIDO EN ACCESIONES DE <i>B. ORELLANA</i>.....	33
3.5 TÉCNICA TETRA PRIMER ARMS-PCR	37
3.5.1 EXTRACCIÓN DE DNA DE ACCESIONES DE <i>B. ORELLANA</i>	37
3.5.2 DISEÑO DE CEBADORES PARA LA TÉCNICA TETRA PRIMER ARMS-PCR.....	38
3.5.3 TETRA PRIMER ARMS-PCR EN TRES ACCESIONES DE <i>B. ORELLANA</i>	39
CAPÍTULO IV.	43
DISCUSIÓN.....	43
4.1 ANÁLISIS DE CALIDAD DE TRANSCRIPTOMAS EN TRES ACCESIONES DE <i>B. ORELLANA</i>.....	43
4.2 LLAMADO Y ANÁLISIS DE SNPs DE TRANSCRIPTOMAS DE TRES ACCESIONES DE <i>B. ORELLANA</i>	43
4.3 SELECCIÓN ALEATORIA Y RELACIÓN FENOTIPO/GENOTIPO DE SNPs.....	45
4.4 ESTABLECIMIENTO DE LA TÉCNICA TETRA PRIMER ARMS-PCR EN <i>B. ORELLANA</i>	47
CAPÍTULO V	49
5.1 CONCLUSIONES	49
5.2 PERSPECTIVAS	51
BIBLIOGRAFÍA.....	52

ÍNDICE DE FIGURAS

Figura 1.1 Distribución de <i>B. orellana</i>	4
Figura 1.2 Morfología de achiote (<i>Bixa orellana</i> L.).....	4
Figura 1.3 Fenotipos de <i>B. orellana</i> representativos.	5
Figura 1.4 Ruta de biosíntesis de carotenoides.....	8
Figura 1.5 Formación de difosfato de isopentenilo por la ruta MEP.....	9
Figura 1.6 Estructuras químicas de los seis carotenoides más frecuentes en la dieta humana.....	10
Figura 1.7 Rutas biosintéticas de los principales apocarotenoides vegetales.....	11
Figura 1.8 Ruta de biosíntesis de bixina.....	13
Figura 1.9 Acumulación de bixina en los diferentes desarrollos de semillas por etapa en tres accesiones de <i>B. orellana</i> ($\alpha=0.05$).....	14
Figura 2.1 Morfotipos de las accesiones de <i>B. orellana</i> estudiadas.	24
Figura 2.2 Accesiones de <i>B. orellana</i> usadas para la validación de los SNPs.....	25
Figura 3.1 Puntuaciones de alineación de transcriptomas de <i>B. orellana</i>	28
Figura 3.2 Distribución en scaffolds de SNP..	30
Figura 3.3 Frecuencia de casos heterocigotos y homocigotos en las accesiones N4, N5 y P12 de <i>B. orellana</i>	31
Figura 3.4 Polimorfismos de un Solo Nucleótido <i>B. orellana</i>	32
Figura 3.5 Localización de los polimorfismos de un solo nucleótido de alta confianza en seis fragmentos obtenidos del genoma de <i>B. orellana</i>	36
Figura 3.6 Posición en scaffolds de SNPs de alta confianza seleccionados aleatoriamente	

de los transcriptomas de *B. orellana* en las accesiones N4, N5 y P12 para el diseño de cebadores..... 37

Figura 3.7 Ilustración esquemática de la técnica Tetra Primer ARMS-PCR... 39

Figura 3.8 Establecimiento de la concentración de cebadores internos y externos durante la estandarización de la técnica Tetra-primer ARMS-PCR utilizando cebadores para SNP I (ID 4591)..... 40

Figura 3.9 Productos de la técnica Tetra-primer ARMS-PCR..... 42

ÍNDICE DE TABLAS

Tabla 2.1 Características fenotípicas de las accesiones de <i>Bixa orellana</i> para el establecimiento de la técnica Tetra Primer ARMS -PCR.	25
Tabla 3.1 Parámetros básicos de calidad realizado por el software FastQC para las secuencias del transcriptoma de alta calidad (Q) de <i>Bixa orellana</i>	27
Tabla 3.2 Resumen del llamado de variantes de <i>B. orellana</i>	29
Tabla 3.3 Tasa de variantes en los transcriptomas de <i>B. orellana</i>	29
Tabla 3.4 Polimorfismos de un Solo Nucleótido seleccionados aleatoriamente de la base de datos de variantes extraídos de los transcriptomas de las accesiones N4, N5 y P12 de <i>B. orellana</i>	33
Tabla 3.5 Polimorfismos de un Solo Nucleótido seleccionados aleatoriamente en los transcriptomas de <i>B. orellana</i> en las accesiones N4, N5 y P12..	35
Tabla 3.6 Concentración de DNA total y relación A260, A280 y A230 nm de seis accesiones de <i>B. orellana</i>	38
Tabla 3.4 Cebadores diseñados para la técnica de Tetra Primer ARMS-PCR.....	41

ABREVIATURAS

SNP: Polimorfismo de un Solo Nucleótido

MAS: Selección Asistida por Marcadores

LCY: Licopeno ciclasa

ABA: Ácido abscísico

IPP: Isopentenil difosfato

MVA: Mevalónica

Acetil-CoA: Acetilcoenzima A

MEP: Metileritritol 4-fosfato

DMAPP: Difosfato de dimetilalilo

G3P: gliceraldehído-3-fosfato

DXP: 1-desoxi-d-xilulosa 5-fosfato

DXPS: 1-desoxi-xilulosa 5-fosfato sintasa

HMBPP: 1-hidroxi-2-metil-2-(E)-butenil 4-fosfato

IDS: IPP/DMAPP sintasa

GGDP: Pirofosfato de geranilgeranilo

PSY: Fitoeno sintasa

CCD: Carotenoides dioxigenasas

ZAS: Zaxinona sintasa

BoLCD: Licopeno dioxigenasa

BoALDH: Aldehído deshidrogenasa

BoMTH: Norbixina metiltransferasa

AP2: Apetala 2

SHP: Shatterproof

SPT: Spatula

DNA: Ácido desoxirribonucleico

RFLP: Polimorfismo de Longitud de Fragmentos de Restricción

RAPD: Amplificación Aleatoria de DNA Polimórfico

SSR: Secuencias Simples Repetidas

SSCP: Polimorfismo de Conformación de una Cadena

PCR: Reacción en cadena de la Polimerasa

PDS: Fitoeno desaturasa

β -LCY: Licopeno β -ciclasa

VCF: Variant Call Format

GATK: Genome Analysis Toolkit

VCF: Variant Call Format

Ts/Tv: Transición/Transversión

ARMS-PCR: Sistema de Mutación Refractario a la Amplificación por Reacción en cadena de la Polimerasa

lncRNA: Ácido ribonucleico largos no codificantes

RESUMEN

Achiote (*Bixa orellana* L.) es una planta leñosa que produce un pigmento llamado bixina, que se acumula principalmente en las semillas. La bixina es importante en la industria alimenticia, cosmética y farmacéutica, sin embargo, aún no se cuenta con variedades agronómicas registradas de achiote para establecer siembras homogéneas que aseguren una producción de semillas y contenido de pigmento predecible. El objetivo de este estudio fue identificar SNP en transcriptomas de tres variantes de achiote para utilizarlos como marcadores moleculares asociados al contenido de bixina, color de la flor y dehiscencia del fruto. Se buscaron SNPs en transcriptomas de tres accesiones morfológicamente contrastantes de achiote (N4, N5 y P12). Se identificaron 17,587 SNPs mediante el programa GATK. La tasa de variantes se obtuvo con ayuda del programa snpEff teniendo 1 variante cada 14,844 bases, una relación $Ts/Tv=1.1497$ y se confirmó que la accesión N4 es heterocigota. Se hizo una selección aleatoria de seis SNPs para el diseño de cebadores para estandarizar la técnica tetra-primer ARMS-PCR con las accesiones N4, N5 y P12 de achiote, conforme a esta prueba, el SNP IX (ID 9086) se relaciona al contenido de bixina y los SNPs I (ID 4591), II (ID 7327), IV (ID 5465) y XV (ID 2875) se relacionan al color de la flor y dehiscencia del fruto. Estos SNPs pueden ser validados para posteriormente usarse como marcadores moleculares para aplicar la Selección Asistida por Marcadores (MAS) en rasgos morfológicos de interés.

ABSTRACT

Achiote (*Bixa orellana* L.) is a woody plant that produces a pigment called bixin, which accumulates mainly in the seeds. Bixin is important in the food, cosmetic, and pharmaceutical industries; however, no registered varieties of achiote exist to establish homogeneous plantings to ensure predictable seed production and pigment content. This study mainly aimed to identify SNPs in transcriptomes of three achiote variants for molecular markers and relate them to bixin content, flower color, and fruit dehiscence. SNPs were searched for in transcriptomes of three morphologically contrasting accessions of achiote (N4, N5, and P12). A total of 17,587 SNPs were identified using the GATK program. The variant rate was obtained with the help of the snpEff program, having 1 variant every 14,844 bases, a ratio $Ts/Tv=1.1497$, and it was confirmed that the N4 accession is heterozygous. A random selection of six SNPs was made, and primers were designed to standardize the tetra-primer ARMS-PCR technique with the N4, N5 and P12 accessions of achiote, according to the results, SNP IX (ID 9086) is related to bixin content and SNPs I (ID 4591), II (ID 7327), IV (ID 5465) and XV (ID 2875) are related to flower color and fruit dehiscence. These SNPs can be validated for later use as molecular markers to apply Marker Assisted Selection (MAS) on morphological traits of interest.

INTRODUCCIÓN

El resurgimiento del uso de colorantes naturales hacen del achiote (*Bixa orellana*), arbusto tropical perenne, una de las fuentes vegetales más interesantes de colorantes naturales (Raddatz-Mota, 2017). *B. orellana* acumula en sus semillas un alto contenido del pigmento apocarotenoide llamado bixina, un pigmento rojo liposoluble (6'-metil-hidrogeno-9'-cis-6,6' diapocaroteno 6, 6'dioato) que representa aproximadamente el 80% del contenido total de carotenoides presentes en el colorante, además, es el único carotenoide que presenta una configuración 9-cis (Smith, 2006; Raddatz-Mota, 2017; Trujillo-Hernández, 2013). La bixina es utilizada por sus cualidades pigmentantes desde la época precolombina hasta la fecha (Rivera-Madrid, 2021), su importancia económica radica en el uso de bixina como colorante en la industria de los alimentos, cosméticos y farmacia (Pinzón *et al.*, 2009; SNICS, 2017).

En vista de su importancia agroindustrial, el grupo de investigación de la Dra. Renata Rivera Madrid, ha invertido esfuerzos en conocer a fondo las enzimas involucradas en la biosíntesis de bixina, los genes que codifican dichas enzimas y la identificación de marcadores moleculares efectivos para la selección de plántulas con características morfológicas de interés agronómico. Anteriormente, se generaron SNPs (Polimorfismos de un Solo Nucleótido) a partir de la variación alélica de genes LCY1 y 2, los cuales participan en la biosíntesis de carotenoides. Estos SNPs separaron en tres grupos de la población analizada por el color de la flor, morfología del fruto y su contenido de bixina (Martín-Rodríguez, 2010; Trujillo-Hernández, 2013; Escobar-Turriza, 2017). Ahora, con la obtención del transcriptoma de *B. orellana*, la búsqueda de marcadores moleculares puede ser aún más amplia y profunda con el objetivo de generar nuevos marcadores moleculares que se relacionen con las características morfológicas de la planta, tales como: el contenido de bixina, número de semillas por capsula, color de la flor, dehiscencia de la cápsula del fruto y color del fruto; este último ha sido correlacionado con la susceptibilidad a plagas y enfermedades (Arce Portuguez, 1984; Duque-Y *et al.*, 2022). Puesto que el valor de *B. orellana* depende de la producción de bixina, es esta la característica que se busca enfatizar al correlacionar los SNPs que se puedan distinguir, sin embargo, los datos que se obtengan podrían ser útiles para enfoques biotecnológicos en la predicción de características deseables, la multiplicación de líneas puras genéticamente y mejoramiento genético.

Por tanto, en este estudio se pretende buscar SNPs en transcriptomas de semilla y hoja pertenecientes a tres variantes de *B. orellana* (N4, N5 y P12) para el establecimiento de la

técnica Tetra Primer ARMS-PCR y asociarlos al contenido de bixina, color de la flor y dehiscencia del fruto.

CAPÍTULO I

ANTECEDENTES

1.1. Generalidades de *Bixa orellana*

1.1.1 Descripción botánica

Bixa orellana L., también conocido como achiote, deriva del vocablo náhuatl 'achiotl', que significa semilla brillante, y el nombre científico de su especie se le dio en honor a su descubridor, Francisco Orellana (Morton, 1960; Raddatz-Mota *et al.*, 2017). Es una planta, tintórea neotropical de interés económico por poseer numerosas propiedades medicinales, nutricionales y económicas. Taxonómicamente, *B. orellana* pertenece a la clase Magnoliopsida, subclase Dilleniidae, orden Malvales, familia Bixaceae y género: *Bixa*; el género *Bixa* contiene cinco especies aceptadas, de las cuales la más popular y explotada comercialmente es *Bixa orellana* L. (Teixeira da Silva *et al.*, 2019; López-Medina *et al.*, 2018).

B. orellana es un arbusto perenne originario de América Central y del Sur tropical, y cultivado por sus semillas en otros países tropicales alrededor del mundo (Figura 1.1) (Teixeira da Silva *et al.*, 2019). Mide de 2 a 5 m (hasta 10 m) de altura, con un diámetro a la altura del pecho de 20 a 30 cm. Copa redondeada y densa. Hojas simples, alternas, grandes y lustrosas, ovadas, de punta larga en el ápice, en pecíolos delgados y largos. Flores grandes, vistosas, dispuestas en corimbos terminales, llevando los pedúnculos de 2 a 4 flores de color rosado o blanco, rojizo o blanco, de 4 a 5.5 cm de diámetro con 5 pétalos, anchos y redondeados. El fruto es una cápsula ovoide a ovoide globosa, pardo-rojiza, de 3 a 5 cm de largo por 3 a 4.5 cm de diámetro, dehiscente por 2 valvas, erizado de pequeños aguijones delgados y blandos hasta de 6 mm; conteniendo aproximadamente de 30 a 45 semillas rojas casi triangulares algo comprimidas y pequeñas, de 5 a 5.5 mm de largo por 4 a 5 mm de ancho, con una testa pulposa de color rojo y un albumen carnoso (Figura 1.2) (Orwa *et al.*, 2009; Sierra-Úlín, 2021; Raddatz-Mota *et al.*, 2017).

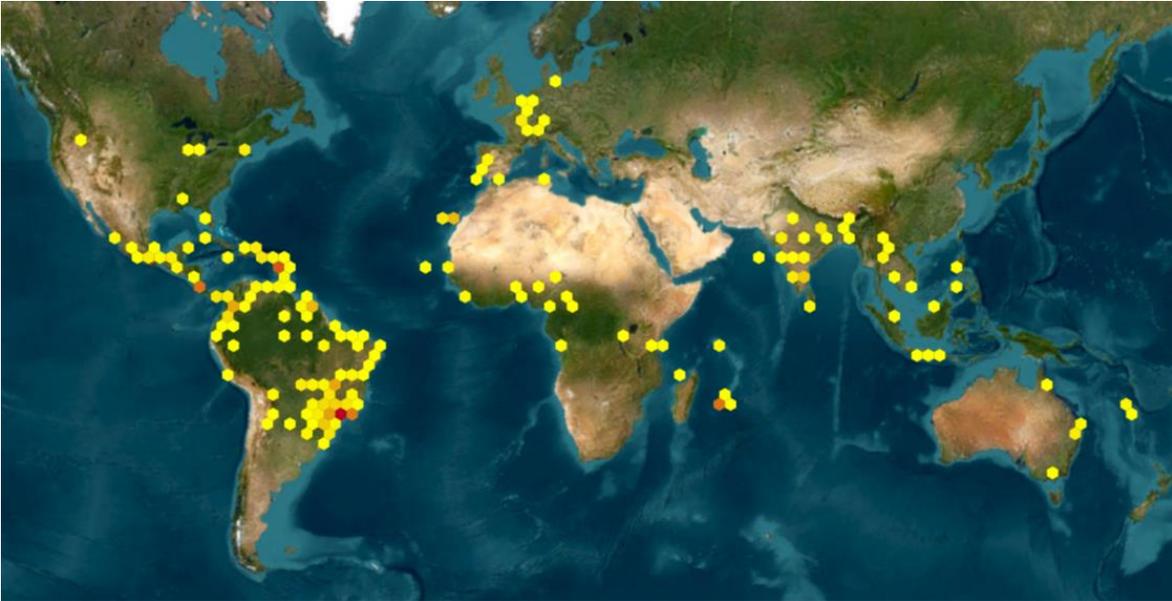


Figura 1.1 Distribución de *B. orellana* (Imagen tomada de P@ntNet).

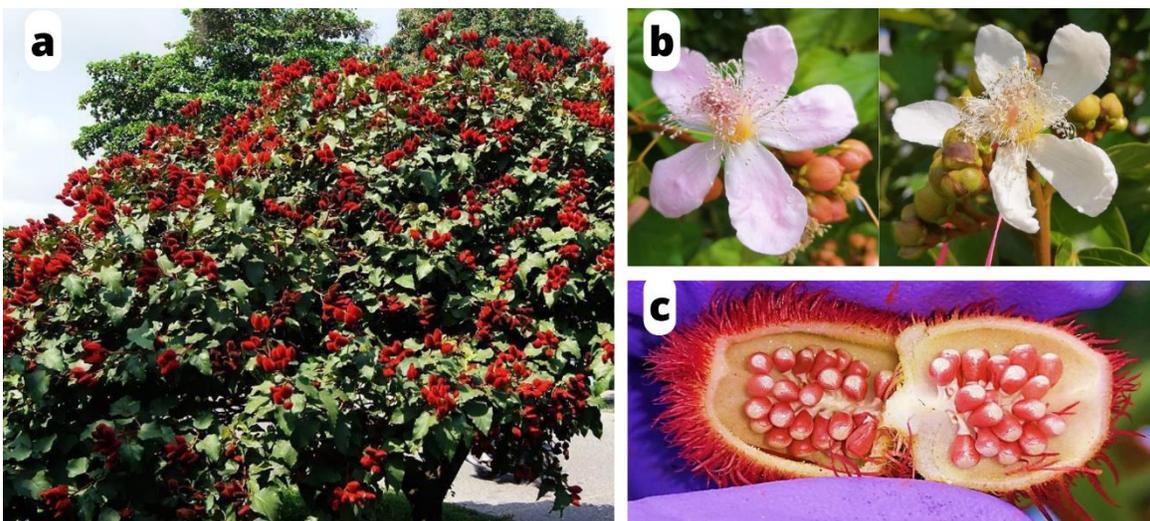


Figura 1.2 Morfología de achiote (*Bixa orellana* L.). a) Planta de *B. orellana*. b) Flores de diferentes variantes de *B. orellana*, presentando los fenotipos de flor rosa y flor blanca. c) Fruto inmaduro de *B. orellana* (Sierra-Ulín, 2021).

En el grupo de investigación de la Dra. Renata Rivera Madrid se han caracterizado tres accesiones de *B. orellana* que se muestrearon en tres localidades distintas, cada una de ellas con características fenotípicas diferentes: grupo 1 con fruto verde y flores blancas;

grupo 2 con fruto verde con espinas rojas y flores rosadas; grupo 3 con frutos rojos con flores moradas (Figura 1.3) (Trujillo-Hernández *et al.*, 2016).

N4



N5



P12



Figura 1.3 Fenotipos de *B. orellana* representativos (Cárdenas-Conejo *et al.*, 2023).

1.1.2 Importancia socioeconómica

B. orellana se considera un árbol multipropósito, pues, aunque comúnmente el colorante que producen las semillas de esta planta ha sido utilizado a nivel mundial en productos alimenticios, su uso se ha extendido a muchos segmentos de la producción industrial. Por lo tanto, ahora se aplica sobre la piel, en forma de maquillaje y protector solar, y hay investigaciones que demuestran que su uso trae beneficios para la salud (Orwa *et al.*, 2009; Vilar *et al.*, 2014). La producción mundial anual de semillas de achiote es de aproximadamente 14,500 toneladas (peso seco). Dos tercios de la producción se comercializan como semillas secas y el resto como colorante. América Latina produce el 60% de la producción mundial total, seguida de África (27%) y Asia (12%). Los principales productores en América Latina son Perú, Brasil y México (Raddatz-Mota *et al.*, 2017; Escobar-Turriza, 2017).

Su importancia económica radica en la presencia del apocarotenoide bixina, el cual representa el 80% del total de los colorantes contenidos en el arilo de la semilla, acompañado de otros compuestos como el norbixin, isobixin y β -caroteno, además, se ha demostrado el aislamiento de numerosos fitoquímicos presentes en *B. orellana* como esteroides, compuestos alifáticos, monoterpenos y sesquiterpenos, flavonoides, polifenoles y taninos, quinonas, terpenoides, alcaloides y triterpenoides (Vilar *et al.*, 2014; López-Medina *et al.*, 2018; Coehlo *et al.*, 2022). En la medicina tradicional tiene numerosas aplicaciones y propiedades farmacológicas como propiedades antibacterianas, antifúngicas, antioxidantes, antiinflamatorias, hemostáticas, de aumento de la motilidad gastrointestinal, anticonvulsivas, analgésicas, antidiarreicas y diuréticas debido a los fitoquímicos presentes que se mencionaron anteriormente (Coehlo *et al.*, 2022)

En la industria alimenticia se utiliza para colorear mantequilla, margarina, mayonesa, salsas, mostaza, chorizo, sopa, jugo, helado, productos de panadería, macarrones y queso. También es ampliamente utilizado en la industria de la impresión y la fabricación de tintes (Vilar *et al.*, 2014)

1.2. Carotenoides

Los carotenoides son pigmentos naturales responsables de los colores que van del amarillo al naranja y al rojo, son sintetizadas por plantas fotosintéticas, algas, bacterias y hongos, pero no por animales (Misawa, 2010; Ellison, 2016). Son un componente esencial de todos los organismos fotosintéticos debido a sus eminentes propiedades, ya que son utilizados

como componentes esenciales del complejo de captación de luz en la fotosíntesis y como precursores de la hormona ABA en las plantas superiores (Lohr, 2009; Misawa, 2010). Los carotenoides dan como resultado órganos de plantas brillantemente pigmentados que atraen pájaros e insectos, lo que ayuda en la dispersión del polen y las semillas y, por lo tanto, ayuda en la polinización y reproducción de las plantas. Por otro lado, dado que los animales no pueden biosintetizar los carotenoides, deben ingerir los pigmentos a su dieta y metabolizarlos para utilizarlos como elementos nutricionales esenciales, por ejemplo, como precursores de la vitamina A, además, se ha demostrado que tienen un papel importante en el mantenimiento de la salud ocular, la función inmunológica y la prevención de enfermedades (Misawa, 2010; Ellison, 2016).

Estos pigmentos pertenecen a los isoprenoides y su estructura básica es C_{40} . Hasta ahora se han identificado más de 700 carotenoides naturales y se pueden distinguir dos tipos de carotenoides: los carotenos (hidrocarburos puros) y las xantofilas (carotenos hidroxilados) (Lohr, 2009; Walter y Strack, 2011). Es debido a la composición de dobles enlaces alternos lo que les permite absorber la luz en el rango visual del espectro (Ellison, 2016).

1.2.1 Biosíntesis de carotenoides en plantas superiores

1.2.1.1 Vías precursoras

En las plantas superiores, los carotenoides se sintetizan y localizan en los plástidos (Figura 1.4); éstos están formados por unidades de isopreno (compuestos de cinco carbonos, IPP) y su biosíntesis se realiza a través de la ruta de los isoprenoides (Raddatz-Mota *et al.*, 2017). Anteriormente se creía que la ruta de MVA era la única forma de sintetizar IPP, se asumió que IPP se sintetiza a partir de acetil-CoA a través del ácido mevalónico, sin embargo, los datos experimentales indicaban constantemente que la ruta de MVA no es exclusiva en la formación de isoprenoides derivados de plástidos. Así la ruta independiente del MVA, conocida como ruta MEP, es una ruta alterna para la formación de IPP y DMAPP y carotenoides (Fraser y Bramley, 2004).

La ruta MEP comprende un total de siete pasos biosintéticos que implican desde la condensación de piruvato con el grupo de aldehído G3P para producir DXP catalizada por la DXPS hasta la conversión de HMBPP en una mezcla 5:1 de IPP y DMAPP con ayuda de la enzima IDS (Figura 1.5) (Fraser y Bramley, 2004; Walter y Strack, 2011).

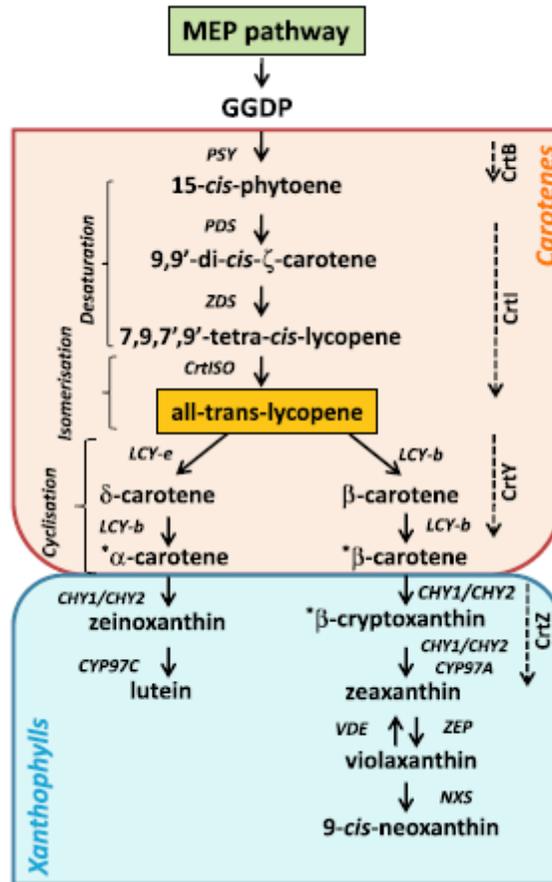


Figura 1.4 Ruta de biosíntesis de carotenoides. GGDP, difosfato de geranylgeranilo; PSY, fitoeno sintasa; PDS, fitoeno desaturasa; ZDS, zcaroteno desaturasa; CRITSO, carotenoide isomerasa; LCY-b, licopeno β -ciclasa; LCY-e, licopeno ϵ -ciclasa; CHY1 y CHY2, b-caroteno hidroxilasas no hemo; CYP97A y CYP97C, citocromo P450 caroteno hidroxilasas que contienen hemo; ZEP, Zeaxantina epoxidasa; VDE, violaxantina de-epoxidasa; NXS, neoxantina sintasa; CrtB, CrtI, CrtY y CrtZ son enzimas multifuncionales en bacterias que catalizan los pasos cubiertos por las flechas discontinuas respectivamente; Todas las enzimas en mayúsculas y cursivas son funcionales en planta; *carotenoides con actividad provitamina A (Ayobamidele, 2018).

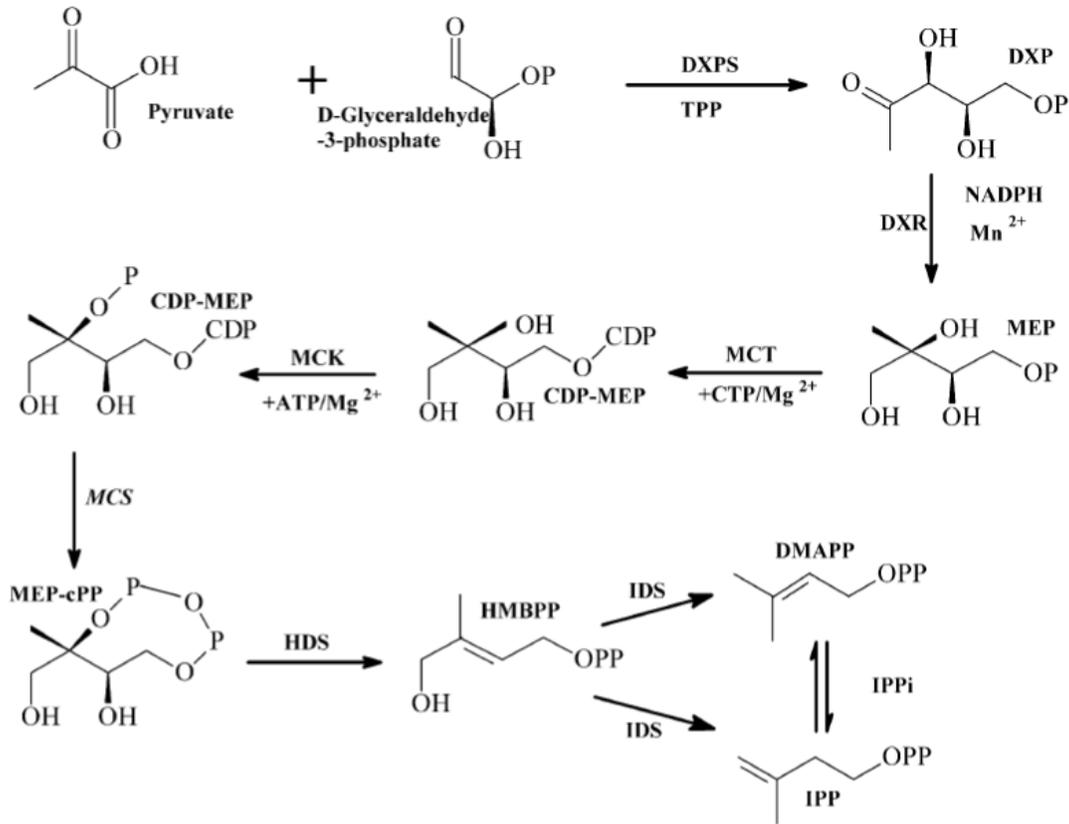


Figura 1.5 Formación de difosfato de isopentenilo por la ruta MEP. TPP, pirofosfato de tiamina; DXP, 5-fosfato de 1-desoxi-d-xilulosa; MEP, 2-C-metil-d-eritritol 4-fosfato; CDP-ME, 4-(citidina 50-difosfo)-2-C-metil-d-eritritol; CDP-MEP, 2-fosfo-4-(citidina 50-difosfo)-2-C-metil-d-eritritol; MEP-cPP, 2-C-metil-d-eritritol-2,4-ciclodifosfato; HMBPP, 1-hidroxi-2-metil-2-(E)-butenil 4-fosfato; IPP, difosfato de isopentenilo; DMAPP, difosfato de dimetilalilo; DXPS, 1-desoxi-d-xilulosa 5-fosfato sintasa; DXR, 1-desoxi-d-xilulosa 5-fosfato reductoisomerasa; MCT, 2-C-metil-d-eritritol 4-fosfato citidil transferasa; CMK, 4-(citidina 50-difosfo)-2-C-metil-d-eritritol quinasa; MCS, 2-C-metil-d-eritritol 2,4-ciclodifosfato sintasa; HDS, 1-hidroxi-2-metil-2-(E)-butenil 4-fosfato sintasa e IPPI, isopentenil difosfato isomerasa (Fraser y Bramley, 2004).

1.2.1.2 Biosíntesis de carotenoides

La gran mayoría de los carotenoides tienen un esqueleto compuesto por 40 átomos de carbono (C₄₀), que pueden clasificarse como tetraterpenos, y se originan a partir del fitoeno, el primer hidrocarburo-carotenoide biosintético que no presenta color. El fitoeno se sintetiza a través de la condensación de dos moléculas de GGDP para producir *cis*-fitoeno (C₄₀) con la eliminación de dos grupos difosfato catalizada por una PSY (Raddatz-Mota *et al.*, 2017; Misawa, 2010). Posteriormente, el *cis*-fitoeno sufre cuatro reacciones de desaturación para la síntesis de licopeno (Fraser y Bramley, 2004). El licopeno es la molécula de carotenoide

más simple (molécula lineal formada por 40 átomos de carbono) y es el sustrato para generar la mayoría de los carotenoides cíclicos a través de la ciclación, la hidroxilación o la oxidación, o la combinación de estos tres tipos de reacción. Estas reacciones tienen lugar en uno de los extremos de la cadena de licopeno; De esta forma, la vía biosintética se ramifica para formar los distintos carotenoides (Figura 1.4 y 1.6) (Raddatz-Mota *et al.*, 2017).

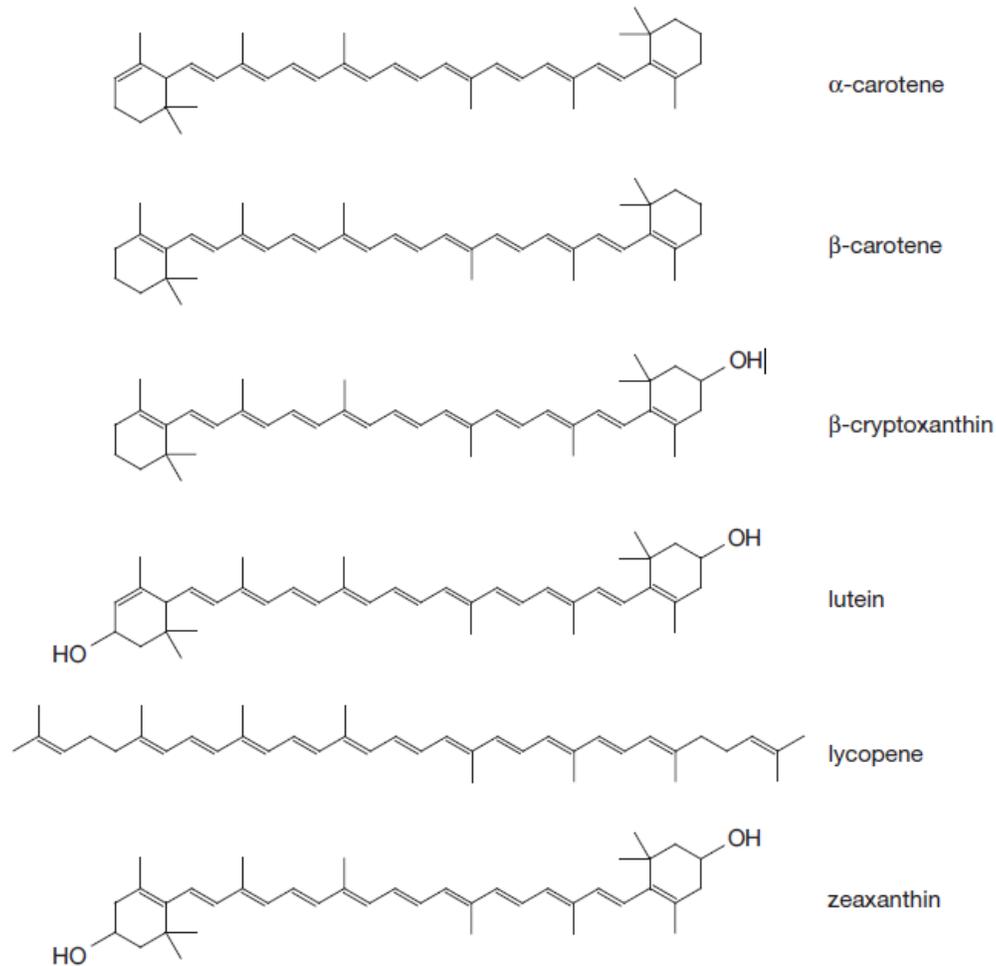


Figura 1.6 Estructuras químicas de los seis carotenoides más frecuentes en la dieta humana. (Ellison, 2016)

1.3. Apocarotenoides

Los apocarotenoides se derivan de la escisión oxidativa de los carotenoides, tienen muchas bioactividades diferentes, algunas de las cuales son esenciales para el metabolismo de las plantas, otras son importantes en la ecología (Cárdenas-Conejo *et al.*, 2015; Walter y Strack, 2011). Si se mantiene un número suficiente de dobles enlaces conjugados en los productos de escisión, pueden tener carácter de pigmento como los carotenoides, también

incluyen compuestos aromáticos volátiles que actúan como repelentes, quimioatrayentes, simuladores de crecimiento e inhibidores, así como las fitohormonas ácido abscísico y estrigolactonas (Walter y Strack, 2011; Beltrán y Barra, 2016). Por tanto, los carotenoides se pueden metabolizar en varios apocarotenoides con funciones biológicas importantes, a través de diferentes procesos, como la escisión enzimática de los carotenoides a través de CCD específicas, pues es debido a su sistema de doble enlace conjugado, rico en electrones, que los carotenoides son susceptibles a la oxidación, provocando la rotura de su columna vertebral y dando lugar a diversos productos de carbonilo (Figura 1.7) (Moreno *et al.*, 2020; Cárdenas-Conejo *et al.*, 2015).

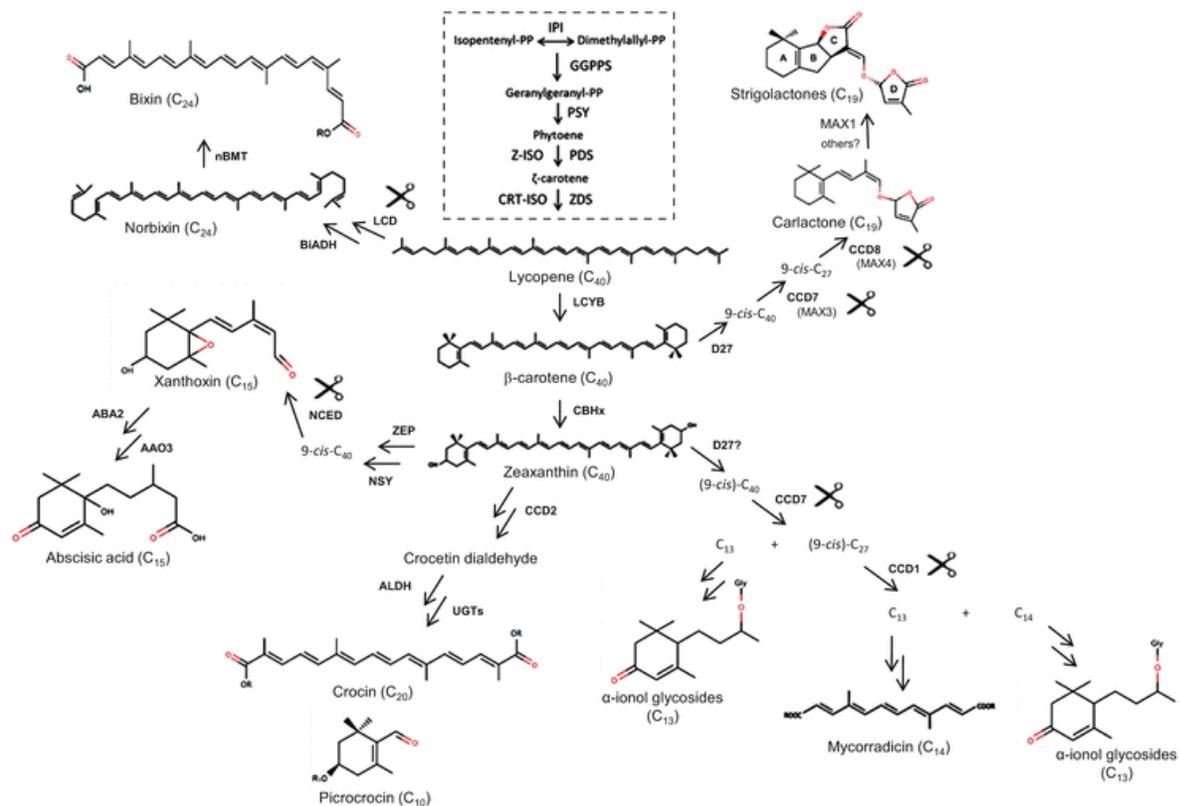


Figura 1.7 Rutas biosintéticas de los principales apocarotenoides vegetales. La vía clásica de los carotenoides con las respectivas enzimas y productos se muestra en un cuadrado discontinuo. LCD dioxigenasa de escisión de licopeno, BiADH bixina aldehído deshidrogenasa, nBMT norbixina metiltransferasa, D27 β-caroteno isomerasa, CCD (1), (2), (7), (8) dioxigenasa de escisión de carotenoides, MAX1 citocromo P450, CBHx caroteno β-hidroxilasa, ALDH aldehído deshidrogenasa, UGT UDPG-glucosiltransferasas, ZEP zeaxantina epoxidasa, NSY neoxantina sintasa, NCED 9-cis dioxigenasas de escisión de epoxicarotenoide, ABA2 xantoxina deshidrogenasa 2, AAO3 aldehído abscísico oxidasa 3 (Moreno-Beltran y Stange, 2016)

1.4. Bixina

Entre los apocarotenoides naturales, la bixina ocupa el segundo lugar en importancia económica. Tiene un color rojo anaranjado y se produce principalmente en las semillas de *B. Orellana* representando más del 80% del total de carotenoides expresados (Us-Camas *et al.*, 2022; Isaza *et al.*, 2019). Bixina (metil hidrógeno 9'-cis-6,6'-diapocaroteno-6,6'-dioato) es uno de los pigmentos más antiguos utilizados por los humanos, ha sido empleado en diversos productos como aditivo de color soluble en la industria alimenticia, cosmética y farmacéutica, debido a que no altera el sabor y no se le han atribuido efectos de tipo genotóxicos o carcinógenos, por tanto, no es solo por la prohibición de colorantes sintetizados químicamente, si no, por las propiedades que bixina tiene, la demanda es cada vez mayor (Bouvier *et al.*, 2003; Isaza *et al.*, 2019; Siva *et al.*, 2010, Agner *et al.*, 2004).

1.4.1 Biosíntesis de bixina

De acuerdo con la propuesta de Cárdenas-Conejo *et al.* (2015), cuando los genes que codifican las enzimas ciclasas regulados a la baja del licopeno tienen poca o nula expresión hay un bloqueo metabólico que impide la síntesis de los carotenoides por debajo del licopeno y provoca una disminución en la concentración de carotenoides cíclicos. De ocurrir este bloqueo, el licopeno podría acumularse aumentando la disponibilidad de este compuesto para la vía de la bixina. La bixina se produce por la escisión oxidativa en los dobles enlaces de licopeno por la acción de las CCD, una gran familia de enzimas clasificadas en seis subfamilias: CCD1, CCD2, CCD4, CCD7, CCD8 y ZAS, que varían en su especificidad de sustrato y sitios de escisión (Us-Camas *et al.*, 2022; Frusciante *et al.*, 2022).

La biosíntesis de bixina comienza con la escisión oxidativa de licopeno en los dobles enlaces 5,6/5',6' por la enzima CCD4 BoLCD, para producir bixina aldehído, a continuación, los grupos aldehído son oxidados a carboxilos por la BoALDH produciendo norbixina. Finalmente, uno de los grupos carboxilo de la norbixina es metilado por la BoMTH para producir bixina (Figura 1.8), (Bouvier *et al.*, 2003; Us-Camas *et al.*, 2022).

En el 2023 Cárdenas-Conejo *et al.* seleccionaron tres accesiones representativas de *B. orellana*, conservadas en el Centro de Investigación Científica de Yucatán (Figura 1.3), anteriormente caracterizadas por Trujillo- Hdz *et al.* (2016) en tres grupos, haplotipos A, B y C. En su estudio, Cárdenas-Conejo *et al.* (2023) evaluaron el contenido de bixina es tres estadios de las semillas, reportando que así como la evidente diferencia morfológica de sus

frutos y flores, el contenido de bixina mostró diferencias significativas, donde en el estadio 3 (S3) la accesión N4 mostró la mayor cantidad de bixina con 154 mg/g D.W (Figura 1.9).

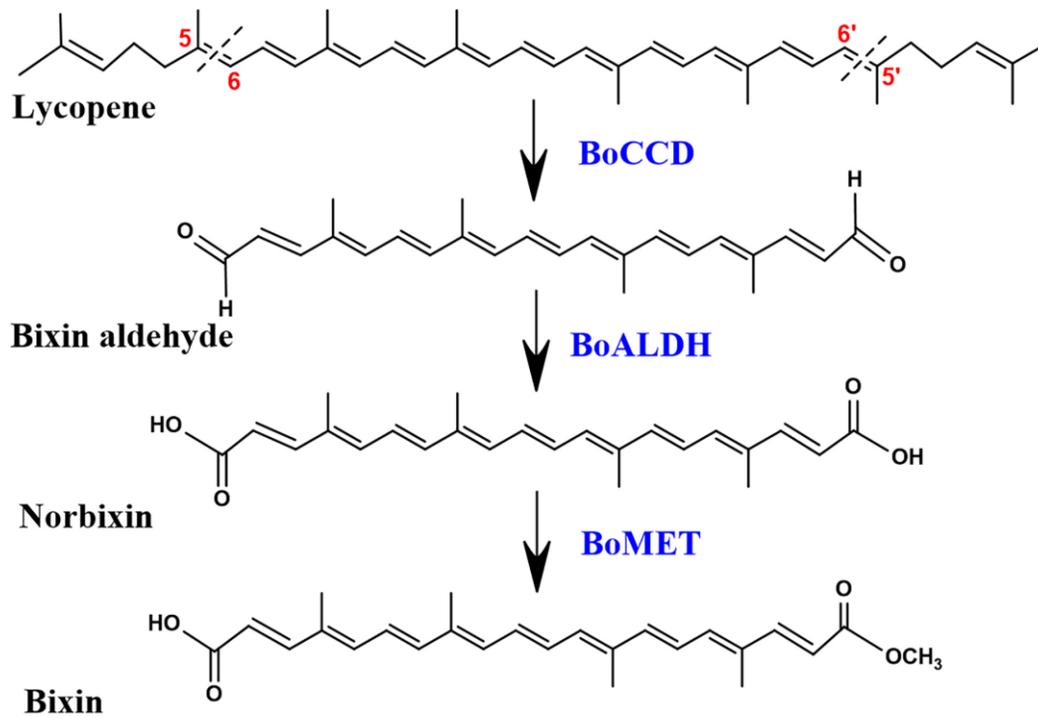


Figura 1.8 Ruta de biosíntesis de bixina. BoCCD, dioxigenasa de escisión de caroteno; BoALDH, bixina aldehído deshidrogenasa; BoMET, norbixina metiltransferasa (Us-Camas *et al.*, 2022).

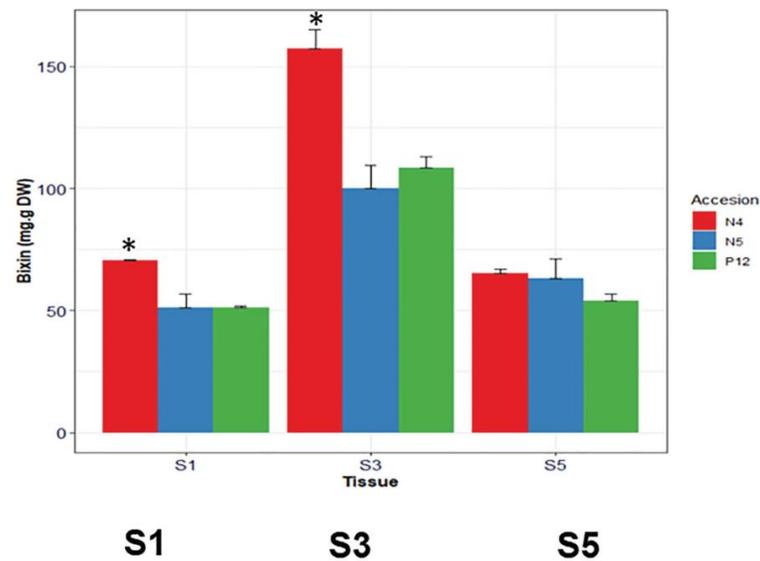


Figura 1.9 Acumulación de bixina en los diferentes desarrollos de semillas por etapa en tres accesiones de *B. orellana* ($\alpha=0.05$) (Cárdenas-Conejo *et al.*, 2023).

1.15. Morfología y genética del fruto

Los frutos de *B. orellana* están formados por valvas dehiscentes o indehiscentes (Tamayo-García *et al.*, 2022) cuya cubierta presenta apéndices espinosos, simples y flexibles, largos o cortos, de acuerdo con la accesión (López *et al.*, 2018).

Las accesiones P12 y N4 (Figura 1.3), preservadas en el Centro de Investigación Científica de Yucatán A.C., presentan morfología contrastante de acuerdo a su dehiscencia, esto a su vez afecta en la producción de bixina, pues bixina es un pigmento sensible a los factores ambientales como la luz, aire y temperatura (Rivera-Madrid *et al.*, 2016). Anteriormente se ha reportado que las accesiones con fruto dehiscente contienen mayor cantidad de bixina que las indehiscentes (Trujillo-Hdz *et al.*, 2016) y en el año 2022 Tamayo-García *et al.* (2022) reportaron que los genes AP2, SHP, and SPT fueron expresados diferencialmente durante seis estadios de desarrollo de los frutos en las accesiones P12 y N4.

No solo se ha reportado la relación de la morfología del fruto con el contenido de bixina, también se ha encontrado una relación con el color de la flor (Arce Portugues, 1984) además, se ha correlacionado el color de la cápsula con la susceptibilidad a plagas y enfermedades (Duque-Y *et al.*, 2022), Arce-Portuguez (1984) menciona que los frutos de

coloración roja y café tienen un mayor grado de susceptibilidad a *Oidium spp.*, por otra parte, frutos de color verde y café son susceptibles al ataque de *Leptoglossus zonatus*.

Finalmente, con la ayuda de herramientas biotecnológicas se podrá predecir características deseables como el contenido de bixina, el número de semillas por cápsula, el número de cápsulas por racimo, el número de racimos por árbol y/o tolerancia a mildu polvoriento (*Oidium spp.*), para la multiplicación de líneas puras genéticamente o para el mejoramiento genético de cultivo del achiote (Avalos, 2013; Quiñones *et al.*, 2014)

1.6. Marcadores moleculares

Existen dos clases de marcadores genéticos: los morfológicos y los moleculares. El uso de marcadores morfológicos en las plantas tiene muchas limitantes, pues su expresión puede estar sujeta a factores ambientales o fenológicos. Dentro de los marcadores moleculares existen dos técnicas que nos permiten conocer cómo se encuentran las proporciones de genes en las poblaciones, estos son de manera indirecta, como los análisis de proteínas, o de manera directa con estudios de DNA (Azofeifa-Delgado, 2006; Rentarías, 2007). Los marcadores moleculares son fenotípicamente neutros, presentan mayor polimorfismo que los morfológicos, pueden ser evaluados desde los primeros estados de desarrollo de las plántulas, son aplicables a cualquier tipo de material vegetal, son independientes de la época del año en que se realiza el análisis, permiten la identificación correcta de la variedad sin necesidad de muchos caracteres y están libres de los efectos epistáticos (Azofeifa-Delgado, 2006; Rentarías, 2007).

Los marcadores moleculares, son herramientas que permiten, a través del análisis del DNA, la asociación fenotipo-genotipo, establecer hipótesis de relaciones filogenéticas entre organismos, estimar la variación dentro de las poblaciones o probar hipótesis de adaptaciones ecológicas utiliza herramientas como los marcadores (Cubero, 2002; Rentarías, 2007). Existen diferentes tipos de métodos en el hallazgo de los marcadores moleculares, y algunos de ellos incluyen el RFLP, RAPD, SSR y SNP (Quintos-Cortes *et al.*, 2021).

1.16.1 Polimorfismo de un solo nucleótido (SNP)

La búsqueda de los marcadores SNP ha ido en aumento dado que los RFLP han ido indicando claramente la existencia de variación sutil generalizada del genoma (Brookes, 1999). Los SNP son marcadores de un solo nucleótido en el material genético, representan el tipo más generalizado de variación de secuencia en los genomas de muchas especies

(regiones codificantes y no codificantes) (Morin *et al.*, 2004). En muchas especies, los SNP ocurren cada 200 a 500 pb (Brookes, 1999; Brumfield *et al.*, 2003).

Han surgido recientemente como marcadores genéticos valiosos para revelar la historia evolutiva de las poblaciones. Gran parte del descubrimiento de SNP se ha realizado *in silico*, lo que significa que la información genómica de múltiples individuos en las bases de datos públicas se analiza para la identificación de polimorfismos putativos. Por otro lado, se han usado métodos de detección alternativos para la detección de SNP, como el SSCP, (Single Strand Conformation Polymorphism), que permite la detección rápida y económica de más muestras, pero con poca sensibilidad, de modo que algunos SNP se pueden perder (Brumfield *et al.*, 2003; Morin *et al.*, 2004).

1.6.2 Antecedentes del uso de marcadores moleculares

En 2021, Bohry *et al.* hicieron un estudio donde identificaron las variantes de DNA en todo el genoma de *Carica papaya* L. entre las líneas élite de Formosa Sekati y JS-12 para predecir sus efectos según la ubicación genómica y desarrollar una lista de genes relacionados con la maduración. Identificaron 28,451 SNPs, 1,061 inserciones y 921 deleciones, de los cuales identificaron un total de 106 genes relacionados con la maduración que estaban asociados con 460 variantes, por lo que se concluyen que pueden convertirse en marcadores de PCR para facilitar el mapeo genético.

El grupo de investigación de la Dra. Renata Rivera Madrid se ha empeñado en la búsqueda de marcadores moleculares tipo SNPs útiles en el mejoramiento genético de *B. orellana*. Martín-Rodríguez (2010) empleó la técnica de SSCP para la obtención de marcadores moleculares específicos de genes involucrados en la síntesis de carotenoides, estos fueron los genes *PDS* y β -*LCY*. Los resultados mostraron que existe variación a nivel de secuencias, donde, el gen β -*LCY* presentó asociación entre el color del fruto y el contenido de bixina en la semilla. Por otro lado, en el gen *PDS* se detectó mayor polimorfismo, sin embargo, no presentó una asociación clara con alguna característica fenotípica. En 2016, Trujillo-Hernández *et al.* estudiaron el gen *BoBLCY1* para demostrar la presencia de marcadores moleculares y asociarlos a las características fenotípicas de 16 variantes de *B. orellana*, logrando identificar dos alelos del gen diferenciados por ocho SNPs, que efectivamente, estaban correlacionados con la morfología y acumulación de bixina de la planta. Posteriormente, Escobar (2017) continuó con la búsqueda de variaciones alélicas pero esta vez en el gen *BoBLCY2* en 28 variantes de *B. orellana*, detectando 35 SNPs

asociados a tres morfotipos con características morfológicas constantes, el mismo autor menciona que la generación de marcadores moleculares nos permiten seleccionar, en una población de achiote, a las variantes que presenten un mayor contenido de bixina en las semillas, contribuyendo de esta manera en los programas de fitomejoramiento de la especie, ya sea para la selección de materiales élite o la detección temprana de plantas potenciales en la producción de bixina.

JUSTIFICACIÓN

La importancia económica del achiote radica en la producción de bixina, un colorante ampliamente usado en las industrias alimentaria, farmacéutica, textil y cosmética, no obstante, otros rasgos morfológicos de *B. orellana* pueden resultar interesantes para su uso en el mejoramiento genético de plantas. Sin embargo, la problemática actual del cultivo del achiote se divide en dos puntos importantes a resolver: a) no se cuenta con variedades registradas que contengan las cualidades agronómicas deseadas para una óptima producción de semilla y contenido de pigmento por kg de semilla; b) a falta de variedades agronómicas no se pueden establecer siembras en campo homogéneas que aseguren una producción de semilla y contenido de pigmento predecible. Una estrategia para abordar la problemática son el uso de herramientas como la bioinformática y biología molecular que permitan resolver ambos problemas identificando variaciones genéticas en el DNA del achiote, tales como los marcadores moleculares, SNPs.

Por tanto, es necesario invertir esfuerzos para su estudio con el propósito de poder predecir, a edad temprana, cuál será su fenotipo y hacer selecciones concretas de plantas de achiote. Con la ayuda de la obtención de SNPs y otras herramientas biotecnológicas es posible seleccionar materiales genéticos con características agronómicas específicas para lograr un mejoramiento genético del achiote más exitoso.

HIPÓTESIS

En *Bixa orellana* existen SNPs asociados al contenido de bixina, por tanto, es posible la presencia de más SNPs relacionados al contenido de bixina, color de flor y dehiscencia del fruto.

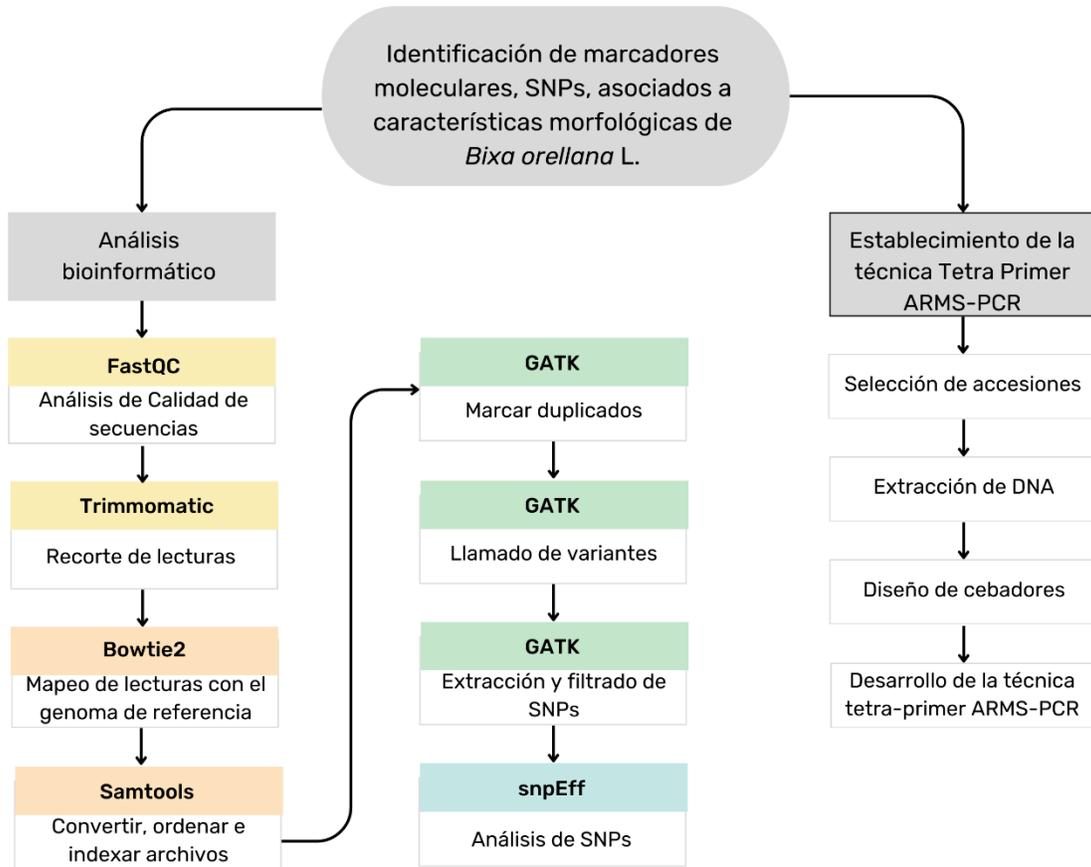
OBJETIVO GENERAL

Identificar SNPs asociados al contenido de bixina, color de la flor y dehiscencia de la cápsula del fruto en tres variantes de *Bixa orellana*.

OBJETIVOS ESPECÍFICOS

- Realizar control de calidad y alineación de lecturas de los transcriptomas de tres variantes de *B. orellana* al genoma de referencia de *B. orellana*.
- Efectuar el llamado de variantes y seleccionar SNPs de interés en los transcriptomas de las tres variantes de *Bixa orellana*.
- Identificar la asociación fenotipo/genotipo con los SNPs seleccionados.
- Establecimiento de la metodología tetra primer ARMS-PCR en tres variantes de *B. orellana*.

ESTRATEGIA EXPERIMENTAL



CAPÍTULO II

MATERIALES Y MÉTODOS

2.1 Análisis bioinformático

2.1.1 Información de los transcriptomas

Los transcriptomas que se usaron para los análisis desarrollados en este proyecto se obtuvieron de la base de datos del laboratorio 27 de la Unidad de Bioquímica y Biología Molecular de Plantas del Centro de Investigaciones Científicas de Yucatán A.C. (CICY), los cuales se construyeron mediante la tecnología de secuenciación Illumina RNA-Seq. Los transcriptomas pertenecen a tejido de hoja joven (H) y semilla en tres estadios de desarrollo (S1, S3 y S5) de las accesiones N4, N5 y P12, en total se analizaron 24 transcriptomas que se obtuvieron de dos individuos por accesión para cada estadio y hoja. La accesión P12 presenta flores blancas y frutos indehiscentes verdes, la accesión N4 muestra flores rosas y frutos dehiscentes verdes con espinas rojas y la accesión N5 presenta flores púrpuras y frutos dehiscentes rojos o amarillos (Figura 2.1).

Los transcriptomas obtenidos en formato fastq fueron sometidos a pasos básicos de limpieza para su posterior análisis bioinformático, es decir, extracción de SNPs. Los siguientes pasos se llevaron a cabo empleando programas bioinformáticos nativos con el sistema operativo Linux/UNIX con los lenguajes de programación Perl y R, este último para realizar las gráficas y posteriores análisis estadísticos.

2.1.2 Análisis de calidad (Q) y limpieza de lecturas

Se utilizó el software FastQC para realizar el control de calidad de los datos sin procesar (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), el programa mostró el puntaje de Phred para cada nucleótido señalando los que estuvieron por debajo del valor $\text{Phred} \geq 20$ y otros aspectos importantes que afectan la calidad (Q) de las lecturas como las secuencias sobrerrepresentadas y el contenido de adaptadores TruSeq Adapter 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTGTAGCCATCTCGTAT-3'.

Con base a los resultados conseguidos con el software FastQC, se recortaron las lecturas con baja calidad con ayuda del software Trimmomatic versión 0.39 (Bolger *et al.*, 2014). Las lecturas se recortaron para quitar nucleótidos de baja calidad, adaptadores y secuencias sobrerrepresentadas, además, se indicó que el tamaño de las lecturas fuese de 35-50 pb.

2.1.3 Mapeo de lecturas

Las secuencias ensambladas del transcriptoma de alta calidad (Q) (con valor Phred ≥ 20) se alinearon y compararon mediante el programa Bowtie2 versión 2.5.0 (Langmead y Salzberg, 2012) con parámetros predeterminados, utilizando el genoma de referencia de *B. orellana*. Para convertir los archivos SAM/BAM, ordenarlos e indexarlos se empleó SAMtools versión 1.6 (Li *et al.*, 2009).

2.1.4 Búsqueda de variantes mediante el método Variant calling (llamado de variantes)

A partir del marcado de duplicados hasta el archivo VCF, las lecturas se procesaron utilizando GATK versión 4.4.0.0. La herramienta *MarkDuplicatesSpark* se utilizó para marcar duplicados. La calidad de las bases se volvió a calibrar con *BaseRecalibrator* y para identificar los supuestos sitios SNP en las transcripciones, se realizó el método llamado de SNP y el filtrado de baja confianza con *haplotypecaller* utilizando parámetros predeterminados. Las gráficas de densidades de SNPs, se realizaron con el software SRPlot (Tang *et al.*, 2023).

2.1.5 Análisis de Polimorfismos de un Solo Nucleótido

La densidad de los SNPs por scaffold se realizó con ayuda del software diseñado por Tang *et al.* (2023), disponible en https://www.bioinformatics.com.cn/plot_basic_SNP_density_by_CMplot_107_en. Se utilizó el programa snpEff v. 5.2c (Cingolani *et al.*, 2012) con el comando -onlyProtein para conocer la frecuencia de casos heterocigotos y homocigotos en las accesiones, así como la relación Ts/Tv y el cambio de bases.

2.2 Establecimiento de la técnica Tetra Primer ARMS-PCR

2.2.1 Material vegetal

Las accesiones N4, N5 y P12 de *B. orellana* fueron usadas para estandarizar el protocolo de la técnica Tetra Primer ARMS-PCR. Los tres individuos tienen características morfológicas contrastantes entre ellas para probar los marcadores SNP significativos (Tabla 2.1). Las semillas de las accesiones seleccionadas fueron colectadas entre 2012-2018, provenientes de la colección del Centro Regional Universitario de la Península de Yucatán (CRUPY) de la Universidad Autónoma de Chapingo (UACH), Temozón Norte, Mérida,

Yucatán. Los datos mostrados en la Tabla 1 fueron determinados por Trujillo-Hdz *et al.* (2016) y Pech-Hoil (2020).

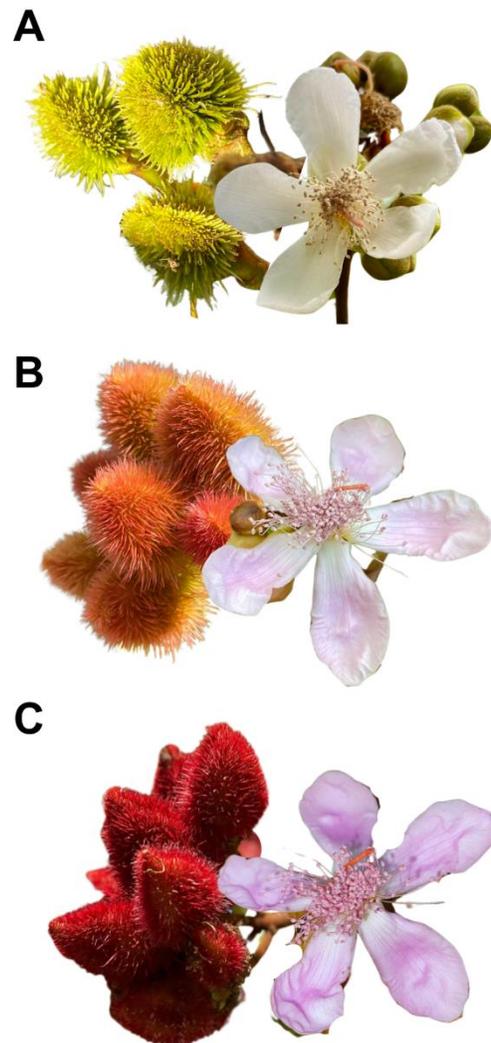


Figura 2.1 Morfotipos de las accesiones de *B. orellana* estudiadas. A) P12, B) N4, C) N5.

Tabla 2.1 Características fenotípicas de las accesiones de *Bixa orellana* para el establecimiento de la técnica Tetra Primer ARMS-PCR.

Accesión	Color de flor	Color de fruto	Dehiscencia
N4	Rosa	Verde/rojo	Sí
N5	Rosa	Rojo	Sí
P12	Blanca	Verde	No

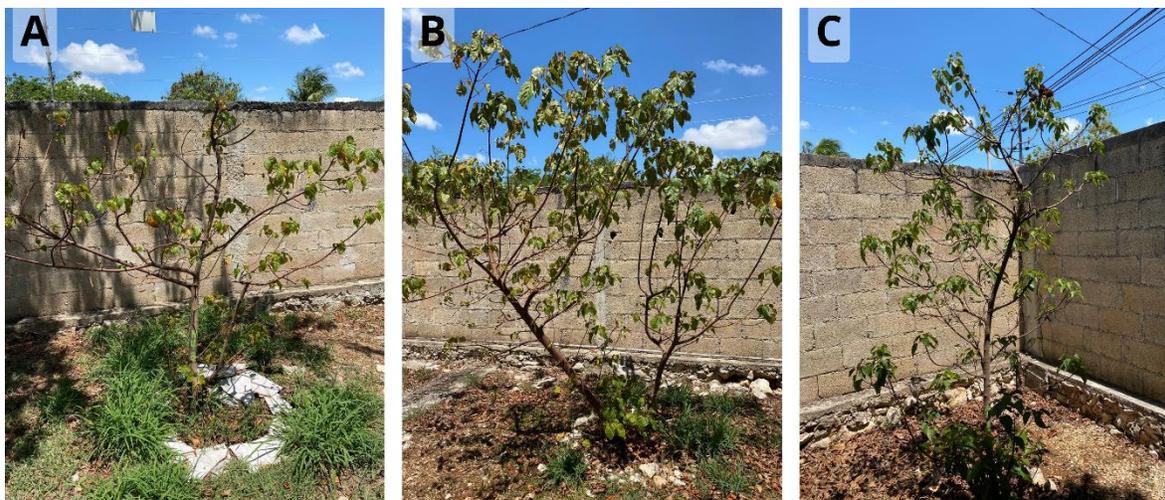


Figura 2.2 Accesiones de *B. orellana* usadas para estandarizar el protocolo de la técnica Tetra Primer ARMS-PCR A) planta madura de la accesión N4, B) planta madura de la accesión N5, C) planta madura de la accesión P12.

2.2.2 Extracción de DNA

La extracción de DNA nuclear se obtuvo a partir de hojas jóvenes tomadas de un sólo individuo para cada accesión, las cuales primeramente se pulverizaron con nitrógeno líquido (NL2) y posteriormente se siguió el protocolo de manufactura del paquete comercial DNeasy Plant (Qiagen Cat. No. 69106) con modificaciones de Valdez-Ojeda *et al.* (2010). La concentración de DNA se evaluó usando un NanoDrop Spectrophotometer ND-1000.

2.2.3 Tetra primer ARMS-PCR

Se estandarizó el método Tetra primer ARMS-PCR, donde, los cebadores específicos y externos a partir de la secuencia que contenga el SNP identificado se diseñaron con el programa desarrollado por Ye *et al.* (2001), utilizando las configuraciones predeterminadas en todos los parámetros. El software informático para diseñar cebadores para

tetracebadores ARMS-PCR está disponible en: <http://primer1.soton.ac.uk/primer1.html>. Las reacciones de PCR se llevaron a cabo en un volumen final de 15 uL, cada reacción contenía 30 ng de DNA molde, tampón PCR 1X (Invitrogen), 0.2 mM dNTP, 1.5 mM de MgCL₂, 5 pmol de cada cebador externo, 75 pmol de cada cebador interno y 1U de la enzima DNA polimerasa (Taq Platinum). Las amplificaciones por PCR se realizaron a 95°C por 5 min, 35 ciclos de 1 min a 95°C, 1 min con la temperatura para hibridación específica de los cebadores internos y externos y 1 min de extensión a 72°C, finalizando con 10 min a 72°C. Los productos de la PCR se separaron mediante electroforesis en gel de agarosa al 1.5% en Tampón TAE 1X que contenía SYBR Safe (Invitrogen Cat. No. S33102) y se corrieron a 85 voltios por una hora.

CAPÍTULO III

RESULTADOS

3.1 Análisis bioinformático

3.1.1 Análisis de calidad y mapeo de lecturas de transcriptomas de tres accesiones de *B. orellana*

El control de calidad de las lecturas secuenciadas se realizó con la herramienta FastQC, posteriormente se realizó un recorte de las secuencias de baja calidad con la herramienta Trimmomatic específicamente a adaptadores y secuencias sobrerrepresentadas en todos los transcriptomas, las secuencias sobrerrepresentadas aparentemente eran colas poli-A. Después de la limpieza, se observó que los puntajes de calidad en escala Phred (Q) (calidad Q) fueron en promedio 40.29, los datos estadísticos observados se han concatenado en la Tabla 3.1.

Tabla 3.1. Parámetros básicos de calidad realizado por el software FastQC para las secuencias del transcriptoma de alta calidad (Q) de *Bixa orellana*.

Accesión	H-N4	H-N5	H-P12	S1-N4	S1-N5	S1-P12	S3-N4	S3-N5	S3-P12	S5-N4	S5-N5	S5-P12
Lecturas crudas												
Total	16039428	19371899	18593551	21099990	19159763	16951315	19950592	14589131	21142742	17960471	18578863	20205633
Contenido GC	40.67	40.67	41.00	41.00	42.33	40.33	42.67	42.33	40.00	43.67	44.33	41.67
Q	40.28	40.31	40.31	39.23	40.29	40.30	40.28	40.31	40.26	40.27	40.31	40.27
Lecturas limpias												
Total	14331764	17412642	16605151	19325662	17911990	15188968	18609979	13473398	18706025	16411420	12414793	17562324
Contenido GC	44.33	44.00	44.33	44.00	44.67	43.67	44.67	44.33	43.67	46.33	46.67	45.67
Q	40.29	40.32	40.31	40.32	40.30	40.29	40.28	40.31	40.25	40.28	40.31	40.27

Posterior al análisis de calidad, se mapearon las lecturas de los transcriptomas de semilla con el genoma de referencia obteniendo para la tasa de alineación general valores entre 86% y 90% a excepción de un transcriptoma, siendo el transcriptoma S5-N4 el que tiene la tasa más alta con 90.1% y el transcriptoma de la accesión N5 en el estadio S5 tiene un 65.18% como tasa de alineación general más baja (Figura 3.1 A). Para los transcriptomas de hoja joven, la tasa de alineación mayor fue observada en la accesión P12

con un valor de 89.2%, seguido de la accesión N4 con 88.6% y finalmente la accesión N5 con 87.6% (Figura 3.1 B).

De igual forma se obtuvieron las tasas de alineación específicas para cada transcriptoma, indicando la frecuencia de lecturas mapeadas al genoma de referencia cero veces, una vez y más de una vez, por supuesto, entre más lecturas alineadas una solo vez es mejor pues indican ser una muestra altamente endógena bien conservada. Afortunadamente la mayoría de los transcriptomas mostraron más del 50% de sus lecturas alineados una sola vez, salvo el transcriptoma S5-N5, mismo que tuvo la tasa de alineación más baja. La fracción de lecturas no mapeadas en el transcriptoma S5-N5 casi iguala a las mapeadas una vez, presentando un valor de 34.8% y 36.8%, respectivamente (Figura 3.1 C).

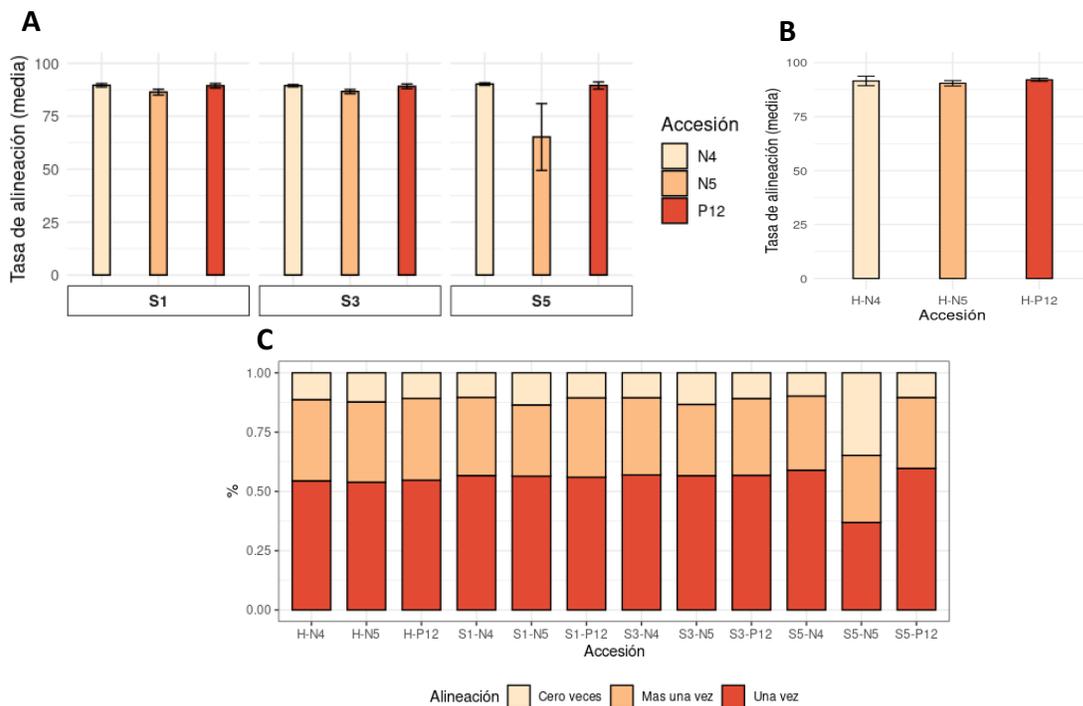


Figura 3.1. Puntuaciones de alineación de transcriptomas de *B. orellana*. A) Tasa de alineación general de los transcriptomas de semillas, B) Tasa de alineación general de los transcriptomas de hoja joven, C) Porcentaje de lecturas emparejadas con el genoma de referencia.

3.2 Llamado de variantes de transcriptomas de tres accesiones de *B. orellana*

Para este estudio sólo se tomaron en cuenta los polimorfismos de un sólo nucleótido, de los cuales se identificaron un total de 17,587 SNP de alta confianza, es decir, las variantes que pasaron el filtro del comando *haplotypecaller* del programa GATK v. 4.4.0.0 y se alojaron en un archivo VCF. Los 17,587 SNPs se encontraron en una escala de 14,844, de acuerdo con el genoma de referencia de *B. orellana*, utilizando transcriptomas de *B. orellana* (Tabla 3.2). Conforme al genoma de referencia, se observó una amplia variación en el número de SNPs por scaffold de acuerdo con los transcriptomas, desde 3,766 SNPs en el scaffold ptg000011l a 47 en el scaffold ptg000020l (Tabla 3.3), y como se visualiza en la Figura 3.2 es debido a la longitud de los scaffolds.

Tabla 3.2. Resumen del llamado de variantes de *B. orellana*.

Número de variantes procesadas	17,587
Longitud total del genoma de referencia	261,065,474
Tasa de variantes de acuerdo con el genoma de referencia	1 variante cada 14,844 bases

Tabla 3.3. Tasa de variantes en los transcriptomas de *B. orellana*.

Scaffold	Longitud	Variantes	Tasa de variantes
ptg000001l	26,012,169	2,554	10,184
ptg000002l	26,251,733	2,198	11,943
ptg000003l	27,395,545	373	73,446
ptg000004l	27,865,768	1,344	20,733
ptg000005l	30,164,334	2,427	12,428
ptg000006l	19,819,941	587	33,764
ptg000009l	29,606,165	2,518	11,757
ptg000010l	12,905,787	683	18,895
ptg000011l	40,842,200	3,766	10,844
ptg000012l	19,716,860	1,090	18,088
ptg000020l	484,972	47	10,318
Total	261,065,474	17,587	14,844

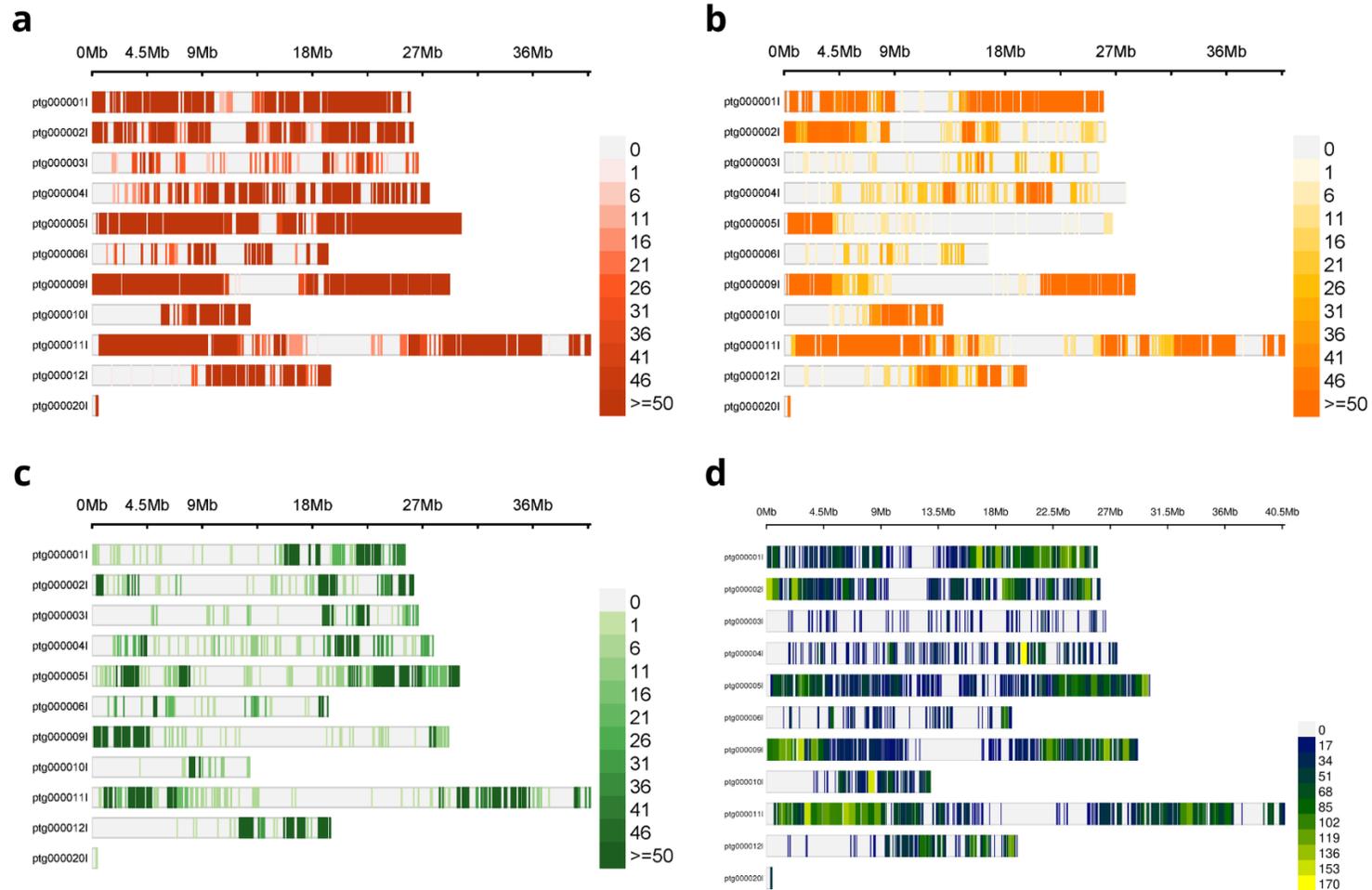


Figura 3.2. Distribución en scaffolds de SNP. a) Densidades en el transcriptoma de *B. orellana* de la accesión N4 en ventanas de 1Mb, b) Densidades en el transcriptoma de *B. orellana* de la accesión N5 en ventanas de 1Mb, c) Densidades en el transcriptoma de *B. orellana* de la accesión P12 en ventanas de 1Mb, d) Densidad total de los transcriptomas de las variantes N4, N5 y P12 de *B. orellana* de 17 K SNP en ventanas de 500 kb. El color de cada rectángulo indica el número de SNP en una ventana particular siguiendo la clave en el lado derecho de la figura.

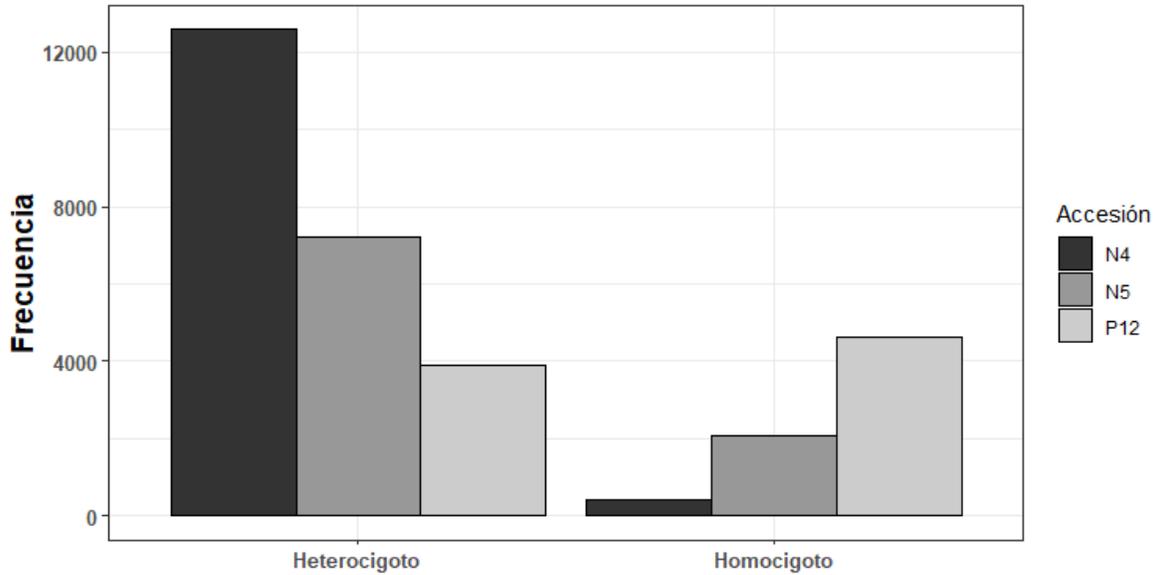


Figura 3.3. Frecuencia de casos heterocigotos y homocigotos en las accesiones N4, N5 y P12 de *B. orellana*.

Alrededor del 55% del total de SNP representaron transiciones con una frecuencia igual a 72,353 y el 45% restante representó transversiones con una frecuencia total de 37,896, teniendo así una relación $Ts/Tv=1.24$. El cambio de bases G/A y C/T fueron los más frecuentes con un recuento de 2,658 y 2,642 respectivamente (Figura 3.2).

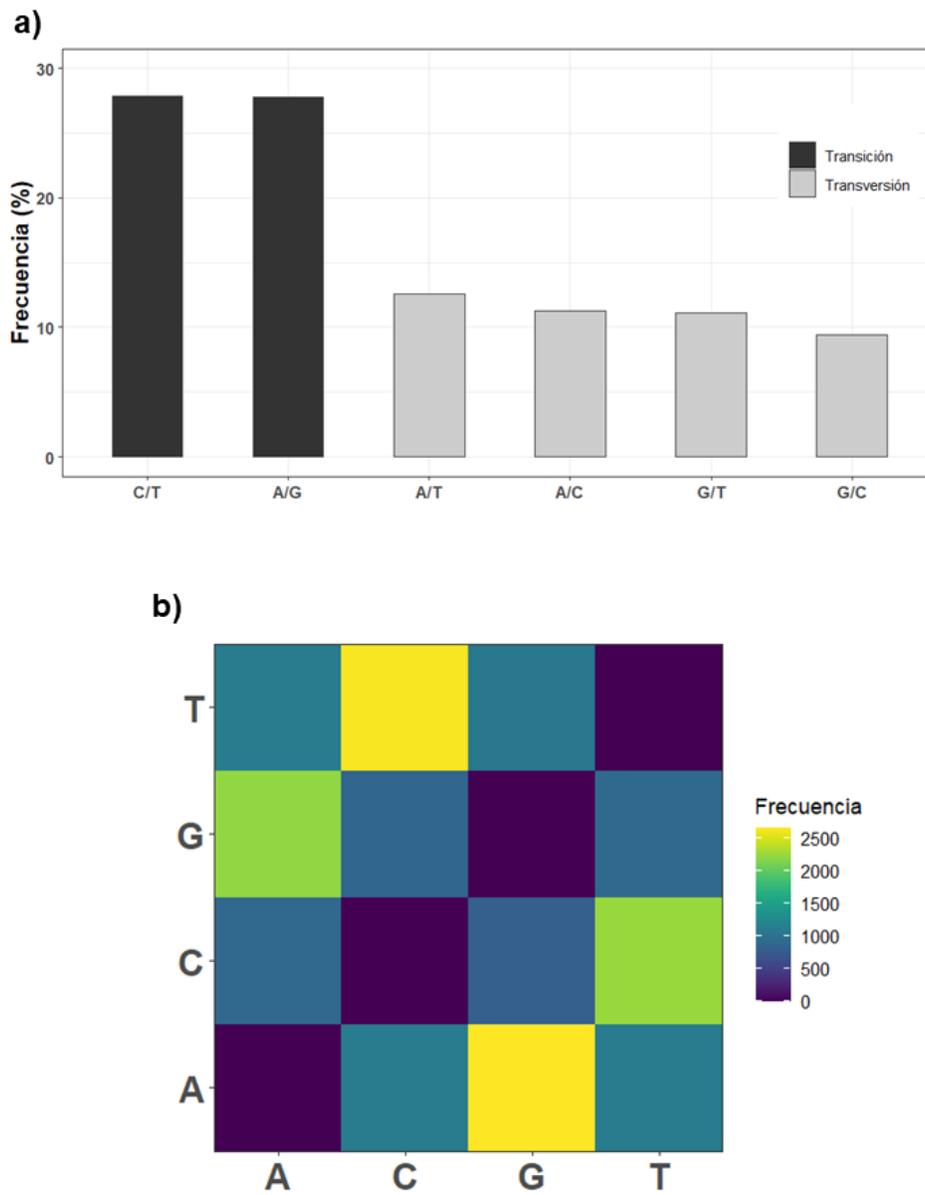


Figura 3.4 Polimorfismos de un solo nucleótido en *B. orellana*. a) Gráfico de barras que muestra la frecuencia de transición y transversión. b) Mapa de calor de las frecuencias del cambio de bases.

3.3 Selección aleatoria y relación genotipo/fenotipo de SNPs en accesiones de *B. orellana*

Se seleccionaron aleatoriamente 20 SNPs de la base de datos de variantes de alta confianza extraídas de los transcriptomas de tres accesiones de *B. orellana*, N4, N5 y P12 (Tabla 3.4). De este conjunto de datos se eligieron seis que cumplieran con una discriminación notoria entre accesiones, tuvieran alta confianza y el contexto en el transcriptoma favoreciera el diseño de cebadores para la técnica Tetra Primer ARMS-PCR.

Tabla 3.4. SNPs seleccionados aleatoriamente de la base de datos de variantes extraídos de los transcriptomas de las accesiones N4, N5 y P12 de *B. orellana*.

ID	Scaffold	Posición	REF	ALT	Calidad
4591	ptg000005l	8735474	G	A	1408.95
7327	ptg000009l	8261880	C	G	1319.95
7885	ptg000009l	23038899	T	C	615.95
5465	ptg000005l	24637143	C	T	702.95
4214	ptg000005l	3657600	T	C	1406.95
9086	ptg000004l	12312894	T	G	998.95
14475	ptg000001l	22080368	G	T	1641.95
5894	ptg000005l	27931710	C	T	738.95
12345	ptg000002l	24547512	C	T	782.95
1435	ptg000011l	7566309	G	T	1024.95
13833	ptg000001l	18198514	T	G	763.95
2838	ptg000011l	32101111	C	T	991.95
10397	ptg000003l	25522091	C	A	527.95
4584	ptg000005l	8647297	A	G	955.95
2875	ptg000011l	32677499	G	T	1149.95
1020	ptg000011l	5745231	A	G	693.95
334	ptg000011l	2496403	T	G	498.95
15649	ptg000006l	18467707	T	C	1055.95
15055	ptg000001l	25012291	G	A	725.95
16220	ptg000012l	13050896	C	T	524.95

De acuerdo con la clasificación dada por Trujillo-Hdz *et al.* (2016), existen tres haplotipos donde se pueden agrupar las accesiones de *B. orellana*, estos son haplotipo A, haplotipo B y haplotipo C, los cuales se encuentran en las accesiones P12, N4 Y N5, respectivamente. Las accesiones pertenecientes al haplotipo A son aquellas que presentan flores blancas y frutos indehiscentes verdes, las accesiones en el haplotipo B muestran flores rosas y frutos dehiscentes verdes con espinas rojas y las accesiones del haplotipo C presentan flores púrpura y frutos dehiscentes rojos o amarillos. Conforme a los seis SNPs seleccionados de los transcriptomas, la accesión N4 contiene dos alelos y las accesiones N5 y P12 muestran un solo alelo (Tabla 3.5). Por tanto, el individuo pertenecientes al haplotipo B (N4) tiene los alelos G/A, C/T, C/G, C/T, T/G y G/T en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente. Del mismo modo, el individuo correspondientes al haplotipo A (P12) tiene el alelo A, T, C, C, G y G en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente. Finalmente, el individuo propio del haplotipo C (N5) tiene los alelos G, C, G, T, T y T en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente (Tabla 3.5, Figura 3.5, Figura 3.6).

Tabla 3.5. SNPs seleccionados aleatoriamente en los transcriptomas de *B. orellana* en las accesiones N4, N5 y P12.

	Scaffold ptg000005l		Scaffold ptg000009l	Scaffold ptg000002l	Scaffold ptg000004l	Scaffold ptg000011l
Posición	G8735474A	C24637143T	C8261880G	C24547512T	T12312894G	G32677499T
	SNP I	SNP IV	SNP II	SNP IX	SNP VI	SNP XV
P12 Haplotipo A 	A	T	C	C	G	G
N4 Haplotipo B 	G/A	C/T	C/G	C/T	T/G	G/T
N5 Haplotipo C 	G	C	G	T	T	T

SNP I

ptg000005l:8735474

	8735473 8735483 8735493 8735503 8735513 8735523 8735533
Reference	TATAATAGAT CG CTGAGTC TTCAACAGG AATATATGTA TTATGTTTCAT CAAGAAATGT ATAGCCTTAG
N4	TATAATAGAT CG CTGAGTC TTCAACAGG AATATATGTA TTATGTTTCAT CAAGAAATGT ATAGCCTTAG
N5	TATAATAGAT CG CTGAGTC TTCAACAGG AATATATGTA TTATGTTTCAT CAAGAAATGT ATAGCCTTAG
P12	TATAATAGAT CG CTGAGTC TTCAACAGG AATATATGTA TTATGTTTCAT CAAGAAATGT ATAGCCTTAG

↓
G/A

SNP II

ptg000009l:8261880

	8261877 8261887 8261897 8261907 8261917 8261927 8261937
Reference	CAGCACTGAC TTG TAATGC ACAAGAAAGA GAAACAGAGA AAGGGGAGCT GGAGTGAGAC TGCCACAGT
N4	CAGCACTGAC TTG TAATGC ACAAGAAAGA GAAACAGAGA AAGGGGAGCT GGAGTGAGAC TGCCACAGT
N5	CAGCACTGAC TTG TAATGC ACAAGAAAGA GAAACAGAGA AAGGGGAGCT GGAGTGAGAC TGCCACAGT
P12	CAGCACTGAC TTG TAATGC ACAAGAAAGA GAAACAGAGA AAGGGGAGCT GGAGTGAGAC TGCCACAGT

↓
G/C

SNP IV

ptg000005l:24637143

	24637141 24637151 24637161 24637171 24637181 24637191 24637201
Reference	AAGTACGTCC T TATTGACA CTTCTCTTCT TCCTCTTGCT TTCTAAAACC TTTAAACCCCT CTTTCGCTGT
N4	AAGTACGTCC T TATTGACA CTTCTCTTCT TCCTCTTGCT TTCTAAAACC TTTAAACCCCT CTTTCGCTGT
N5	AAGTACGTCC T TATTGACA CTTCTCTTCT TCCTCTTGCT TTCTAAAACC TTTAAACCCCT CTTTCGCTGT
P12	AAGTACGTCC T TATTGACA CTTCTCTTCT TCCTCTTGCT TTCTAAAACC TTTAAACCCCT CTTTCGCTGT

↓
C/T

SNP VI

ptg000004l:12312894

	12312893 12312903 12312913 12312923 12312933 12312943 12312953
Reference	TCATGTTTAG T TTTCTTTCT TTCTTCCTCA TTGCTTTTGG GTCGTGAATG TCCAGAAATG ATGGTTGAAA
N4	TCATGTTTAG T TTTCTTTCT TTCTTCCTCA TTGCTTTTGG GTCGTGAATG TCCAGAAATG ATGGTTGAAA
N5	TCATGTTTAG T TTTCTTTCT TTCTTCCTCA TTGCTTTTGG GTCGTGAATG TCCAGAAATG ATGGTTGAAA
P12	TCATGTTTAG T TTTCTTTCT TTCTTCCTCA TTGCTTTTGG GTCGTGAATG TCCAGAAATG ATGGTTGAAA

↓
T/G

SNP IX

ptg000002l:24547512

	24547511 24547521 24547531 24547541 24547551 24547561 24547571
Reference	ATTCAACCGG C TTTACTCT CTTGGAACAA GAACGAACAT ATTCACCCAA TAGACATTGC ACAATTGGAC
N4	ATTCAACCGG C TTTACTCT CTTGGAACAA GAACGAACAT ATTCACCCAA TAGACATTGC ACAATTGGAC
N5	ATTCAACCGG C TTTACTCT CTTGGAACAA GAACGAACAT ATTCACCCAA TAGACATTGC ACAATTGGAC
P12	ATTCAACCGG C TTTACTCT CTTGGAACAA GAACGAACAT ATTCACCCAA TAGACATTGC ACAATTGGAC

↓
C/T

SNP XV

ptg000011l:32677499

	32677498 32677508 32677518 32677528 32677538 32677548 32677558
Reference	AAGGAAGAAG G GAACAGCCA CAAATAAAGA CTAATATCT GTGTGAACAT TTAGATAATT TTGGGTTTGG
N4	AAGGAAGAAG G GAACAGCCA CAAATAAAGA CTAATATCT GTGTGAACAT TTAGATAATT TTGGGTTTGG
N5	AAGGAAGAAG G GAACAGCCA CAAATAAAGA CTAATATCT GTGTGAACAT TTAGATAATT TTGGGTTTGG
P12	AAGGAAGAAG G GAACAGCCA CAAATAAAGA CTAATATCT GTGTGAACAT TTAGATAATT TTGGGTTTGG

↓
G/T

Figura 3.5. Localización de los SNPs de alta confianza en seis fragmentos de los transcriptomas de *B. orellana*. Los SNPs seleccionados están marcados en color rojo con relleno amarillo, mostrando el cambio de nucleótido señalado por una flecha.

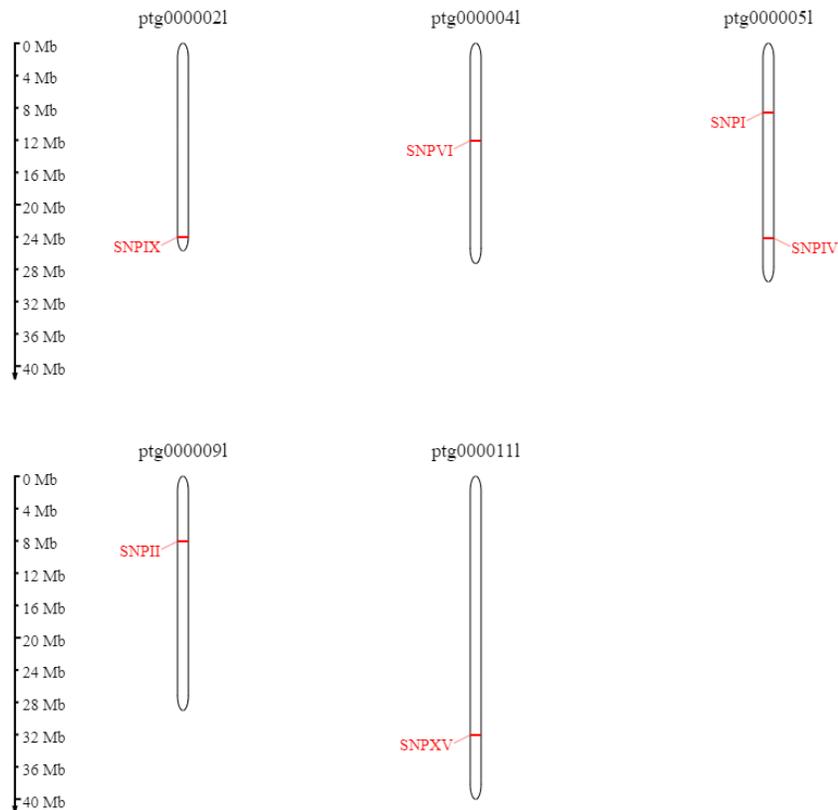


Figura 3.6. Posición en scaffolds de SNPs de alta confianza seleccionados aleatoriamente de los transcriptomas de *B. orellana* en las accesiones N4, N5 y P12 para el diseño de cebadores.

3.5 Técnica Tetra Primer ARMS-PCR

3.5.1 Extracción de DNA de accesiones de *B. orellana*

Se extrajo el DNA nuclear de seis accesiones de *B. orellana* con características morfológicas contrastantes descritas en la Tabla 2.1 La concentración de DNA nuclear se estimó utilizando un NanoDrop Spectrophotometer ND-1000 y los resultados están concatenados en la Tabla 3.6.

Tabla 3.6 Concentración de DNA total y relación A260, A280 y A230 nm de tres accesiones de *B. orellana*.

Accesión	ng/μl	260/280	260/230
N4	25.57	1.90	1.76
N5	26.11	1.89	1.68
P12	22.20	1.88	1.65

3.5.2 Diseño de cebadores para la técnica Tetra Primer ARMS-PCR

La técnica Tetra Primer ARMS-PCR (Amplification Refractory Mutation System–Polymerase Chain) utiliza cuatro cebadores (dos cebadores externos y dos cebadores internos) en una única reacción PCR seguida de la separación en gel de agarosa. La combinación de los cebadores externos amplifica el fragmento más grande que comprende el SNP, se muestra en todos los genotipos y actúa como control positivo; mientras se produce el fragmento de cebadores externos, este sirve como plantilla para los cebadores internos que son para alelos específicos, generando fragmentos de diferentes tamaños que son fácilmente discriminativos en un gel de electroforesis (Figura 3.7) (Ye *et al.*, 2001).

Para esta técnica, se seleccionaron seis polimorfismos de la lista de la Tabla 3.4. los cuales debían cumplir con una discriminación notoria entre accesiones, tener puntuaciones de alta confianza y que su posición en los scaffolds favoreciera el diseño de cebadores para la técnica Tetra Primer ARMS-PCR. Una vez seleccionados se procedió al diseño de los cebadores para seis polimorfismos, los cuales se encuentran enlistados en la Tabla 3.7.

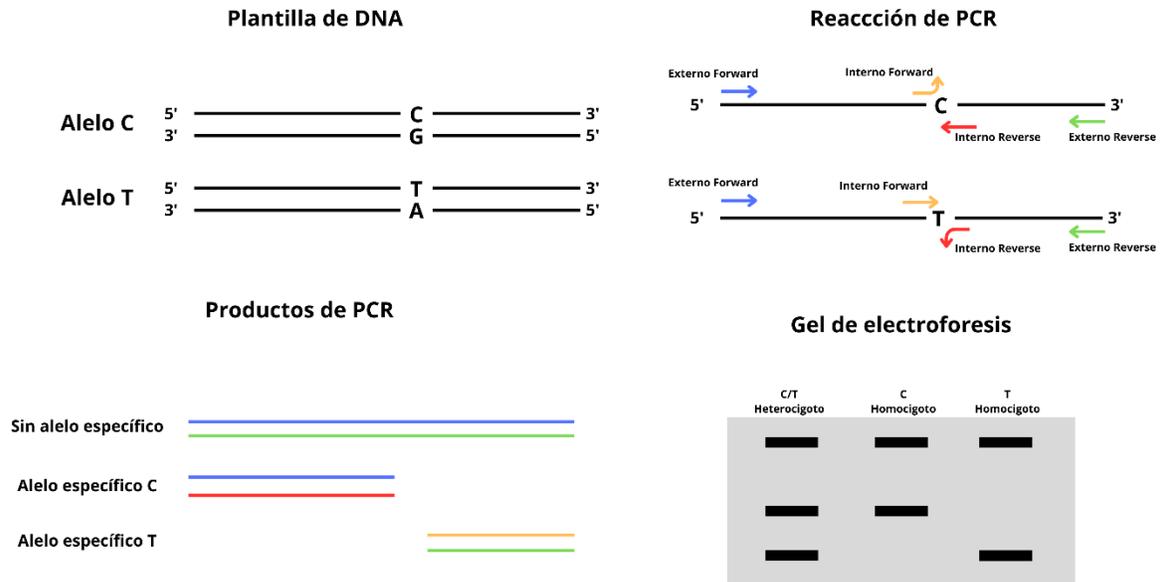


Figura 3.7. Ilustración esquemática de la técnica Tetra Primer ARMS-PCR.

3.5.3 Tetra Primer ARMS-PCR en tres accesiones de *B. orellana*

Una vez que se dispuso de los cebadores se procedió a la amplificación del DNA previamente aislado de las diferentes variantes, tomándose un individuo por variante para las pruebas de amplificación por PCR. Así para ir probando la metodología sugerida por Ye *et al.* (2001) se probaron dos concentraciones de cebadores externos e internos, el primero con 5 pmol de cada cebador (externo e interno) y el segundo con 5 pmol de cada cebador externo y 75 pmol de cada cebador interno. La Figura 3.9 presenta que la reacción con la concentración de 5 pmol de cebadores externo y 75 pmol de cebadores internos fue la que mostró las bandas ligeramente más intensas para los amplicones con los cebadores internos, sin embargo, la banda amplificada con los cebadores externos se vio afectada. Finalmente, se optó por usar la concentración de 5 pmol de cebadores externo y 75 pmol de cebadores internos de cebadores para probar la técnica de tetracebadores ARMS-PCR en los SNPs enlistados en la Tabla 3.5.

La figura Figura 3.9 mostró diferencias entre las tres accesiones confirmando la detección de los polimorfismos, especialmente con el SNP IX (ID 9086) (Figura 3.9E) donde tres bandas de 298, 207 y 144 pb se amplificaron mostrando a la accesión N4 como heterocigota, mientras que los homocigotos se mostraron en las accesiones N5 y P12 al

mostrar dos bandas a 298 y 207 pb para el alelo T, y a 298 y 144 pb para el alelo C, respectivamente. Los productos de la técnica Tetra-primer ARMS-PCR usando los cebadores para el resto de los SNPs mostraron a las tres accesiones como homocigotas debido a que solo se observan dos bandas amplificadas. Los cebadores del SNP I (ID 4591) (Figura 3.9A) amplificaron dos bandas de 229 y 166 pb en las accesiones N4 y N5 correspondiente al alelo G y dos bandas de 229 y 118 pb en la accesión P12 confirmando el alelo A. Los cebadores del SNP II (ID 7327) (Figura 3.9B) amplificaron dos bandas de 309 y 173 pb para las accesiones N4 y N5, y dos bandas de 309 y 192 pb para la accesión P12, correspondientes los alelos G y C, respectivamente. Para los cebadores del SNP IV (ID 5465) (Figura 3.9C) se amplificaron dos bandas de 206 y 164 pb en las accesiones N4 y N5 conforme al alelo C y dos bandas de 306 y 198 pb para la accesión P12 que tiene el alelo T. Los cebadores del SNP VI (ID 9086) (Figura 3.9D), amplificaron dos bandas de 367 y 248 para la accesión N4 que corresponde al alelo T, la accesión N5 presenta una banda de 367 pb indicando que no hubo amplificación de los cebadores internos y la accesión P12 no presentó amplificación. Finalmente, los cebadores del SNP XV (ID 2875) (Figura 3.9F) amplificaron dos bandas de 382 y 210 pb en las accesiones N4 y N5, y dos bandas de 382 y 231 pb para la accesión P12, correspondientes a los alelos T y G, respectivamente.

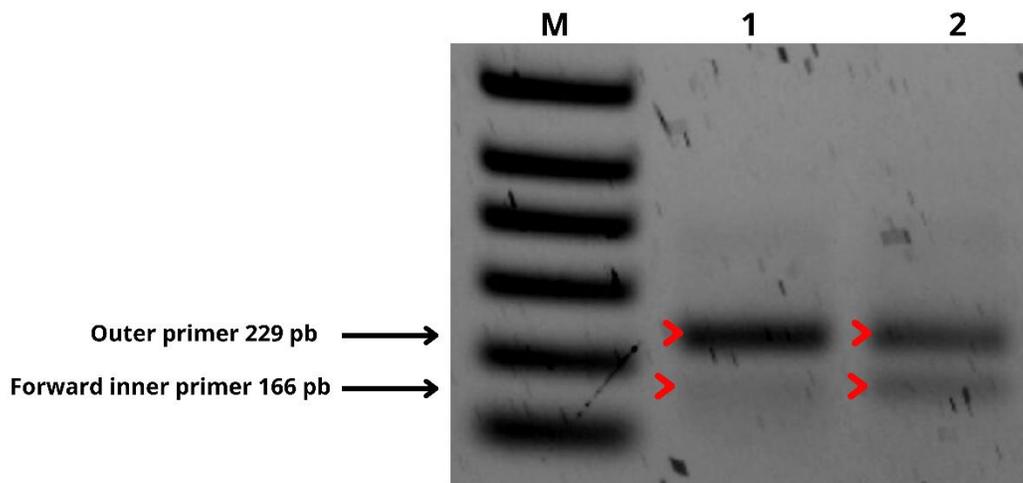


Figura 3.8. Establecimiento de la concentración de cebadores internos y externos durante la estandarización de la técnica Tetra-primer ARMS-PCR utilizando cebadores para SNP I (ID 4591). Electroforesis en gel de agarosa al 1.5%. M: marcador molecular (1Kb). 1: accesión N4 de *B. orellana* con 5 pmol de cada cebador (externo e interno). 2: accesión N4 de *B. orellana* con 5 pmol de cada cebador externo y 75 pmol de cada cebador interno.

Tabla 3.7 Cebadores diseñados para la técnica de Tetra Primer ARMS-PCR.

SNP	ID	Cebadores	Secuencia (5' - 3')	Alelo	Tm (°C)	Longitud (pb)
I	4591	Forward inner 1	AATTAGCTGAGATAAGAATATAATAG CTG	G	56	166
		Reverse inner 1	TATATTCCTGTTGAAAGACTCAG AGT	A	56	118
		Forward outer 1	TTAATTCATGAAAACATACTTTTATCC		56	229
		Reverse outer 1	ATTCATCTCTCTAGGTTTAAATTGAT		56	
II	7327	Forward inner 2	AGGAAATAAACTTGCAGCACTGAC ATC	C	65	192
		Reverse inner 2	TCTCTGTTTCTTTCTTGTGCATT TCC	G	65	173
		Forward outer 2	AGGAACCAAATGTCACATTTGCTATTTG		65	309
		Reverse outer 2	AAATTCTGCTCCGTACAGCTGATAAGAC		65	
IV	5465	Forward inner 4	CATCTTCTCGATCCAAAAGTACG TATT	T	65	198
		Reverse inner 4	GCAAGAGGAAGAAGAGAAGTGTCA AGAG	C	65	164
		Forward outer 4	ATAGATTCCAGAGCACAGACGTCTGAA		65	306
		Reverse outer 4	GAATTTAACCAACAAAGAGGGGAGAAAA		65	
VI	9086	Forward inner 6	TTTTTATCAAATTTGTCATGTTT GGG	G	63	174
		Reverse inner 6	AAAAGCAATGAGGAAGAAAGAAAG GAA	T	63	248
		Forward outer 6	CTCCAAAACCAAAAATAAACACTCAAATAC		63	367
		Reverse outer 6	CTTGAAGAAGAGACAGCTTGATGTGTAG		63	
IX	12345	Forward inner 9	AGAAGCCAATTGCAAATTCAAC CTGT	T	66	207
		Reverse inner 9	TTCGTTCTTGTCCAAGAGAGTA ACGG	C	66	144
		Forward outer 9	AGGTAAACAGATGCAATTGACAGCAAAG		66	298
		Reverse outer 9	CTGATTGCCGAGACTTGATAATCACATT		66	
XV	2875	Forward inner 15	GATAAATGGATATAACTGCAAGGAAG AGGT	T	63	210
		Reverse inner 15	GATATTTAGTCTTTATTTGTGGCTGT CCC	G	63	231
		Forward outer 15	GAGCCCTACTTAATGATAATTGAAAGGG		63	382
		Reverse outer 15	GTCTCTTACGAGAAGTTTCGACAAATTG		63	

Tm Temperatura de fusión

*Las discrepancias de los cebadores específicos de alelo están resaltadas en negrita

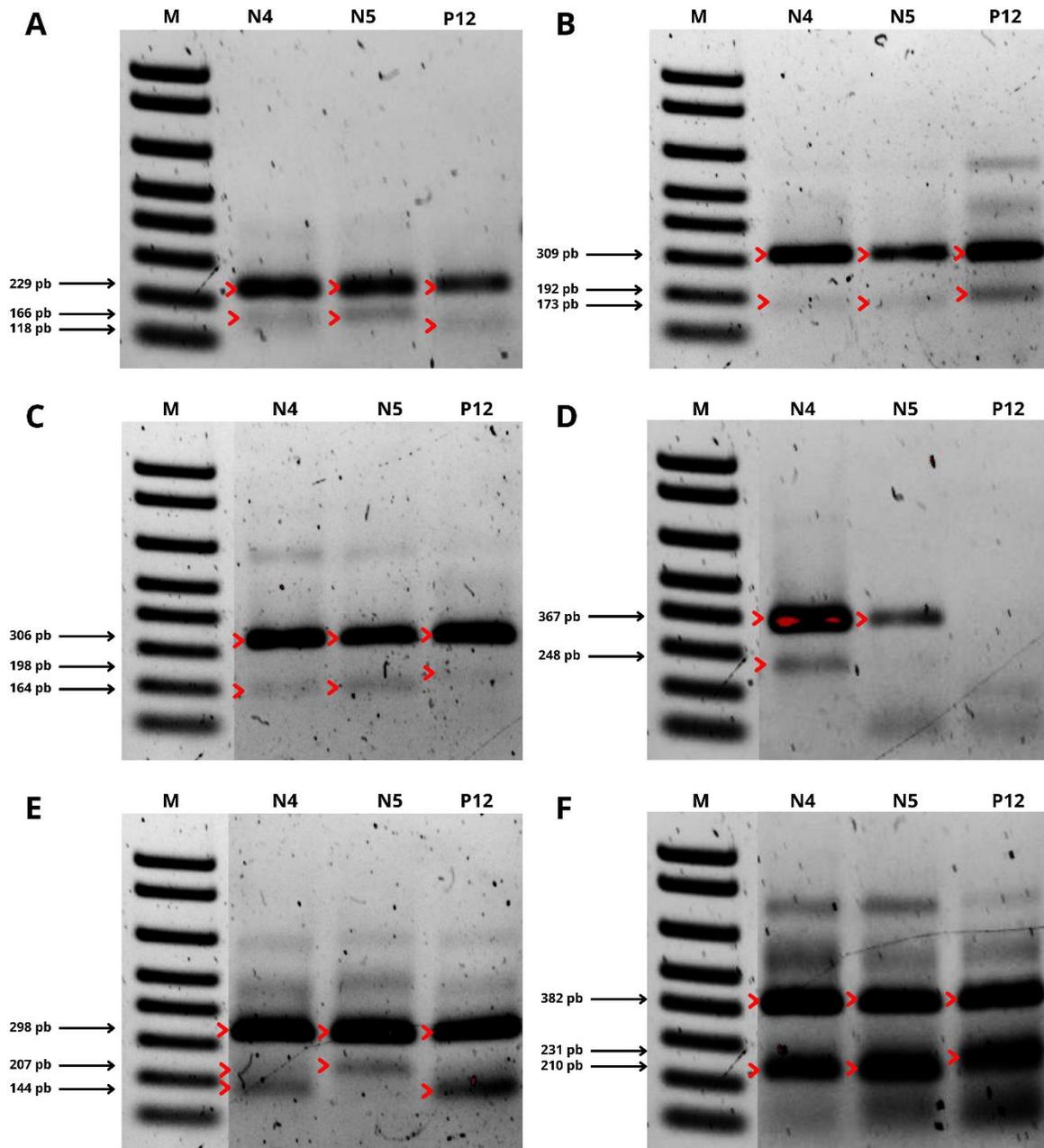


Figura 3.9. Productos de la técnica Tetra-primer ARMS-PCR utilizando cebadores para A) SNP I (ID 4591), B) SNP II (ID 7327), C) SNP IV (ID 5465), D) SNP VI (ID 9086), E) SNP IX (ID 12345) y F) SNP XV (ID 2875). Electroforesis en gel de agarosa al 1.5%. M: marcador molecular (1Kb). N4: accesión N4 de *B. orellana*. N5: accesión N5 de *B. orellana*. P12: accesión P12 de *B. orellana*.

CAPÍTULO IV

DISCUSIÓN

4.1 Análisis de calidad de transcriptomas en tres accesiones de *B. orellana*

El achiote es importante por el uso de la bixina como colorante en la industria alimenticia, cosmética y farmacéutica, por esto en México se ha prestado atención a su cultivo, por lo que se ha incluido en programas de bienestar social (SNICS, 2017). Sin embargo, aún no se cuenta con variedades con las cualidades agronómicas deseadas para el establecimiento de siembras en campo homogéneas que aseguren una producción de semilla y contenido de pigmento predecible. En respuesta a ello, las herramientas de bioinformática y biología molecular permiten resolver la problemática actual identificando variaciones genéticas en transcritos del achiote, tales como los marcadores moleculares que estamos desarrollando y que es necesario validar.

Para la obtención de los transcriptomas que se analizaron, se seleccionaron tres accesiones representativas, P12, N4 y N5, de los grupos de haplotipos A, B y C, respectivamente, previamente caracterizados por Trujillo-Hdz *et al.* (2016). Esta selección fue debido a su contraste morfológico, pero sobre todo por el contenido de bixina, aspecto de mayor interés en la planta, donde la accesión N4 muestra el mayor contenido de bixina alcanzando 154 mg/g P.S. durante el estadio de desarrollo de la semilla 3 (S3) (Cárdenas-Conejo *et al.*, 2023). Asimismo, de la accesión N4 se obtuvo el genoma de referencia para el mapeo de los transcriptomas (Figura 3.1).

En este estudio, se obtuvo en promedio 16.6 Mb y 16.1 Mb de lecturas limpias de semilla y hoja en tres accesiones de achiote, respectivamente, donde se observó que los puntajes de calidad (Q) fueron en promedio 40.29 (Tabla 3.1).

4.2 Llamado y análisis de SNPs de transcriptomas de tres accesiones de *B. orellana*

Los SNPs son una variación abundante del genoma y causantes de la diversidad genética entre los individuos de una especie; puede ocurrir con diferentes frecuencias en diferentes especies a lo largo de todo el genoma (Brookes, 1999; Morgil *et al.*, 2020) como sucede en los estudios realizados en los genomas de *Populus trichocarpa* (Geraldés *et al.*, 2013) y *Carica papaya* (Bohry *et al.*, 2021) donde se identificaron 34,131 y 28,451 SNPs. Esto mismo sucede en estudios realizados con transcriptomas, donde en 2 individuos *Allium cepa* (líneas OH1 and 5225) se reportaron 65.675 SNPs en dos transcriptomas de aproximadamente 500 Mb (Duangjit *et al.*, 2013), en ocho individuos de *Pisum sativum* se

identificaron 35,455 SNPs en transcriptomas con longitudes de 138 a 205 Mb (Duarte *et al.*, 2014), en *Avena sativa* se identificaron 6,634 SNPs en diez individuos concernientes a diez variedades donde los transcriptomas tenían una longitud promedio igual a 137 Mb (Kim *et al.*, 2022) y en este estudio se identificaron 17,587 SNPs en transcriptomas de tres individuos de *B. orellana* (Tabla 3.2).

Los 17,587 SNPs identificados en el transcriptoma de *B. orellana* se encontraban distribuidos irregularmente en los once segmentos divididos del transcriptoma (Figura 3.2 d), observándose una amplia variación en el número de SNPs por scaffold que van de 3,766 SNP en el scaffold ptg0000111 a 47 en el scaffold ptg0000201 (Figura 3.2 d, Tabla 3.3). Aunque la longitud de los diferentes scaffolds es variable, la distribución inconstante de los SNP a través de transcriptomas de las especies es frecuente y esperado (Brookes, 1999; Geraldine *et al.*, 2013; Morgil *et al.*, 2020)

En el transcriptoma de *B. orellana* en promedio se observaron 1 SNP cada 14.8 Kb, una tasa relativamente alta en comparación con el estudio realizado por Riju y Arunachalam, (2010), quienes en transcriptomas de cinco accesiones de *Theobroma cacao* reportaron una tasa igual 1 SNP cada 0.17 Kb.

La figura 3.2 a-c muestra los SNPs de alta confianza de las accesiones N4, N5 y P12, de los cuales se extrajeron 26,520, 10,042 y 5,871, respectivamente. A pesar de que el genoma de referencia pertenece a la accesión N4, en la gráfica de densidades de SNPs de la accesión N4 se puede observar una elevada frecuencia de estos en comparación con la densidad de SNP de las accesiones N5 y P12 (Figura 3.2 a-c), esto debido a su heterocidad mencionada por primera vez por Trujillo-Hdz *et al.* (2016), mismo que se comprueba con la frecuencia de alelos heterocigotos en la Figura 3.3, pues la suma de los alelos heterocigotos en la accesión N4 supera por mucho la frecuencia de alelos heterocigotos de las accesiones N5 y P12, con este resultado se sustenta aún más el argumento de que la accesión N4 es una accesión heterocigota.

La relación transición a transversión (Ts/Tv) es un parámetro que se puede utilizar para inferir sitios de puntos críticos mutacionales, para reconstrucción de árboles filogenéticos y la estimación de la divergencia (Javaid *et al.*, 2022; Wang *et al.* 2015). Aunque algunos estudios hacen mención de relaciones bajas o altas de Ts/Tv (Nosková *et al.*, 2023), no existen referencias que sustenten esta clasificación, por tanto, es más frecuente señalar si el valor es mayor o menor que 1 y así inferir si existen más transiciones o transversiones.

En los transcriptomas de *B. orellana* se obtuvo una relación Ts/Tv de 1.24 (Figura 3.4). La relación Ts/Tv mayor que 1 en *B. orellana*, indica que las variaciones genéticas están compuestas mayormente por transiciones en lugar de transversiones (Figura 3.4), algo que es común de observar (Riju y Arunachalam, 2010). Sin embargo, tal no es el caso en el transcriptoma de *T. cacao*, con Ts/Tv=0.75 (Riju y Arunachalam, 2010), especie más cercana filogenéticamente a *B. orellana* (Cárdenas-Conejo *et al.*, 2015). Se ha reportado que en el genoma humano los exones están enriquecidos por Ts y se deduce que este patrón es debido a que es menos probable que estas variaciones genéticas den lugar a la sustitución de aminoácidos (Guo *et al.*, 2017). Por otro lado, los cambios estructurales inducidos por SNPs en las regiones no codificantes del genoma son difíciles de capturar, ya que no codifican ni afectan la secuencia de la proteína, sin embargo, hoy en día con los análisis del genoma completo, no solo se estudian los efectos que tienen los Ts y Tv en los exones, también podemos conocer su efecto en zonas no codificantes del genoma, que a su vez pueden alterar la estructura de lncRNA, influir en la afinidad de los factores de transcripción y que en consecuencia modifican la actividad del elemento regulado (Singh y Kumar, 2023; Chekanova, 2015; Guo *et al.*, 2017). Por tanto, en *B. orellana*, es muy probable que, la relación Ts/Tv se vea afectada si se hace un llamado de SNPs usando el genoma completo, además de que alguno de los SNPs localizados en las zonas no codificantes esté estrechamente relacionado con la regulación de alguno de los genes involucrados en la ruta de carotenoides o bixina y éste afecte directamente en la acumulación del pigmento bixina.

4.3 Selección aleatoria y relación fenotipo/genotipo de SNPs

Se seleccionaron aleatoriamente 20 SNPs, de manera que se tuviera una muestra representativa de los SNPs extraídos en transcriptomas. Con base en los 20 SNPs extraídos aleatoriamente, se seleccionaron seis que cumplieran con lo siguiente: 1) Discriminación notoria entre accesiones, es decir, que cada accesión tenga un alelo diferente, o si es el caso, que tenga dos alelos, de este modo será más fácil distinguir los alelos en un gel de agarosa al usar la técnica de Tetra Primer ARMS-PCR (Ye *et al.*, 2001); 2) Alta confianza. El valor observado en los archivos vcf que indica llamadas de alta confianza (Cosgun y Oh, 2020); 3) Posición en el genoma que favorezca el diseño de cebadores, tomando en cuenta el %GC en la región (Medrano y de Oliveira, 2014) y que la temperatura de fusión (T_m) entre los cuatro cebadores a usar por reacción de PCR fuese

igual o muy cercana, de acuerdo con el software informático desarrollado por Ye *et al.* (2001) para diseñar los tetracebadores.

El SNP 4214 (Tabla 3.4) es una variante que no fue considerada para usarla en la validación de SNPs, pues a pesar de que tiene una de las puntuaciones de calidad más alta (QUAL=1406.95), la región donde se encuentra presenta alto contenido de GC (%GC) y conforme a las indicaciones de Medrano y de Oliveira (2014), no es conveniente utilizar la técnica de Tetra Primer ARMS-PCR con el SNP 4214, dado que se tienen que romper las fuertes estructuras secundarias que se forman por los enlaces de hidrógeno entre citosina y guanina, por lo que se necesitarían altas temperaturas para poder romper estos enlaces dificultando la amplificación correcta de la región deseada. Así, se seleccionaron seis SNPs que cumplieran con las características explicadas anteriormente, los cuales se nombraron SNPI, SNPII, SNPIV, SNPVI, SNPIX y SNPXV con ID 4591, 7327, 5465, 9086, 12345 y 2875, respectivamente (Tabla 3.5, Figura 3.10).

Al examinar la alta heterogocidad en la accesión N4 (Figura 3.3) (Trujillo-Hdz *et al.*, 2016), era de esperarse que en la selección aleatoria de SNPs presentará mayormente casos con dos alelos, y como se observa en los seis SNPs seleccionados para validar las variantes, es la accesión N4 la única que dispone de dos alelos (Tabla 3.5). Esta característica se emplea para relacionar los seis SNPs seleccionados con la morfología de las accesiones, por ejemplo, el SNPI (ID 4591) puede ilustrar que cuando se tiene el alelo A esta accesión pertenece al haplotipo A, si están presentes los alelos G/A esta accesión pertenece al haplotipo B y si se tiene el alelo G pertenece al haplotipo C (Trujillo-Hdz *et al.*, 2016).

A pesar de que se carece de un análisis estadístico debido a la falta de accesiones secuenciadas y analizadas, los resultados obtenidos en este estudio no son descartables pues muestran un panorama de la distribución de SNPs en las accesiones de *B. orellana* y al mismo tiempo se pueden relacionar con la morfología de las accesiones (contenido de bixina, color de la flor o del fruto y/o dehiscencia del fruto). Estas relaciones genotipo/fenotipo usando SNPs sin modelo lineal pueden ser validadas y reproducibles tal como lo demuestra el trabajo realizado por Yang *et al.* (2023), donde buscaban autenticar a *Panax ginseng* usando un SNP localizado en la región intergénica trnL-trnF, utilizaron la técnica Tetra Primer ARMS-PCR para validar el SNP logrando la confirmación de la especie objetivo y la detección de adulteración en una sola prueba. Otro ejemplo es el estudio realizado por Garcés-Claver *et al.* (2007), quienes asociaron un SNP a la pungencia

(fenotipos picantes y no picantes) de *Capsicum spp*, eligiendo la técnica Tetra Primer ARMS-PCR para su validación debido a su fácil manejo y bajo costo.

4.4 Establecimiento de la técnica Tetra Primer ARMS-PCR en *B. orellana*

Aunque la importancia de achiote radica en el uso de bixina como colorante, esta no es la única característica de interés, algunos aspectos como el color de la flor y la morfología del fruto son importantes al momento de seleccionar una accesión de achiote, por ejemplo, es más fácil identificar una accesión por el color de la flor y esta al mismo tiempo se relacione al contenido de bixina. Existen estudios que han confirmado la relación de características morfológicas con aspectos de interés agronómico, como Trujillo-Hdz *et al.* (2016), quienes mencionan que la dehiscencia del fruto está relacionada al contenido de bixina, ó Duque-Y *et al.* (2022) y Arce-Portuguez (1984) que observan la correlación del color del fruto con la resistencia a plagas y enfermedades. Con las bases de datos morfológicos que se tienen de las tres accesiones, y ahora con la obtención de los SNPs en cada accesión, no solo se pueden correlacionar con el contenido de bixina, también se puede hacer con otro aspecto morfológico de interés. Sin embargo, para poder hacer esta correlación reproducible es necesario validar el SNP.

Tetra-primer ARMS–PCR es considerado un método rápido, robusto, de bajo costo y fácil de usar para la validación de SNPs (Ye *et al.*, 2001; Chiapparino *et al.*, 2004). Pese a que existen otras técnicas, como marcadores CAPS, que se apoya de enzimas de restricción, se prefiere el tetra-primer ARMS–PCR por su rapidez y selectividad, disminuyendo los falsos positivos por los dos desajustes en el extremo 3' desestabilizando el emparejamiento de bases entre los cebadores y el DNA molde (Tabla 3.7) (Ye *et al.*, 2001; Garcés-Claver *et al.*, 2007; Park *et al.*, 2020). Es por estas razones que se ha optado y se recomienda usar la técnica tetra-primer ARMS–PCR para validar los seis SNPs seleccionados.

En este estudio, se estableció del análisis ARMS-PCR con tetracebadores en SNPs de *B. orellana*. Se eligieron las accesiones N4, N5 y P12 para poder distinguir la presencia o ausencia de los alelos localizados bioinformáticamente mediante electroforesis. Para optimizar la técnica ARMS-PCR con tetracebadores se debe tener en cuenta la intensidad del fragmento amplificado que está determinado por la temperatura de hibridación, la calidad de DNA y la concentración de los cebadores externos e internos (Ye *et al.*, 2001, Medrano y Oliveira, 2014). Medrano y Oliveira (2014) mencionan que cambios de solo 0.07

pmol de los cebadores externos e internos son suficientes para modificar de manera evidente su intensidad o inferir en su formación, en este trabajo, la modificación en la concentración de los cebadores internos alteró ligeramente la intensidad de la banda, sin embargo, Medrano y Oliveira (2014) mencionan que estas modificaciones pueden afectar directamente al resto de los amplicones, tal como se observa en la Figura 3.9.

La prueba con los tetracebadores para el SNP IX (ID 9086) (Figura 3.9E) ubicado en el scaffold ptg000002l: 24547512, fue la que mostró mayor discriminación entre accesiones, pues cada accesión tenía un patrón diferente en la posición y número de bandas en el gel de agarosa, además, en *B. orellana*, la presencia de tres bandas es el indicador de su heterogocidad, genotipo que anteriormente se ha relacionado al contenido de bixina siendo la accesión N4 la que tiene mayor contenido de bixina (Trujillo-Hdz *et al.*, 2016; Cárdenas-Conejo *et al.*, 2023).

A pesar de que no se observó la presencia de tres bandas en los geles de agarosa para la accesión N4 en los análisis de tetracebadores ARMS-PCR para los SNPs I (ID 4591), II (ID 7327), IV (ID 5465) y XV (ID 2875), las amplificaciones son similares a la accesión N5, por lo que, estos SNPs pueden relacionarse con el color de flor rosa y los frutos dehiscentes, morfología que ambas accesiones presentan (Tabla 2.1). Por otro lado, la accesión P12 que tiene flores blancas y fruto indehiscente presentó un patrón diferente en la prueba de tetracebadores para los mismos SNPs, diferenciándolo de las accesiones N4 y N5 (Figura 3.9 A, B, C, F).

El SNP VI (ID 9086) en la prueba de tetracebadores ARMS-PCR fue el único que no se pudo comparar entre accesiones, pues solo la accesión N4 amplificó los cebadores internos. La accesión N5 solo presenta la amplificación de los cebadores externos, banda que se usa como control positivo (Ye *et al.*, 2001), por tanto, se puede concluir que la accesión N5 no posee el SNP identificado bioinformáticamente (Tabla 3.5). La accesión P12, en contraste, no presentó amplificación de ningún cebador por lo que se le atribuye a la calidad del DNA (Ye *et al.*, 2001) (Figura 3.9 D).

Los resultados en este estudio son prometedores debido a la alta confianza de los SNPs extraídos, por tanto, en un futuro se pretende validar los SNPs con la técnica tetra-primer ARMS-PCR, especialmente con los cebadores para el SNP IX (ID 9086), y usar los SNPs validados como marcadores moleculares para una selección precisa accesiones con características específicas y con el tiempo proceder al mejoramiento genético.

CAPÍTULO V

5.1 CONCLUSIONES

Se extrajeron un total de 17,587 SNPs de los transcriptomas de *B. orellana* con una amplia variación en el número de SNPs por scaffold, desde 3,766 SNPs en el scaffold ptg000011l a 47 en el scaffold ptg000020l y una tasa de variantes igual a 1 variante cada 14,844 bases. Además, cerca del 55% del total de SNPs representaron transiciones con una frecuencia igual a 72,353 y el 45% restante representó transversiones con una frecuencia total de 37,896, donde el cambio de bases G/A y C/T fueron los más frecuentes con un recuento de 2,658 y 2,642 respectivamente, de estos datos se estimó una relación $Ts/Tv=1.24$.

Se seleccionaron 20 SNPs al azar, de los cuales se eligieron seis que cumplieran con una discriminación notoria entre accesiones, tuvieran alta confianza y que su posición en los scaffolds favoreciera el diseño de cebadores para la técnica Tetra Primer ARMS-PCR. De acuerdo con los análisis bioinformáticos la accesión P12 tiene el alelo A, T, C, C, G y G en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente. La accesión N4 tiene los alelos G/A, C/T, C/G, C/T, T/G y G/T en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente. Finalmente, la accesión N5 tiene los alelos G, C, G, T, T y T en las posiciones ptg000005l: 8735474, ptg000005l: 24637143, ptg000009l: 8261880, ptg000002l: 24547512, ptg000004l: 12312894 y ptg000011l: 32677499, respectivamente.

Se usaron los SNPs anteriormente descritos para la prueba Tetra Primer ARMS-PCR en las accesiones N4, N5 y P12. La prueba realizada en el SNP IX (ID 9086) mostró diferencias entre las accesiones en el número y posición de las bandas amplificadas, donde la accesión N4 presentó tres bandas de 298, 207 y 144 pb demostrando su heterogocidad con la presencia de los alelos T y C, mientras que las accesiones N5 y P12 se observaron como homocigotas con dos bandas a 298 y 207 pb para el alelo T, y a 298 y 144 pb para el alelo C, respectivamente. Por tanto, el SNP IX se relaciona con el contenido de bixina de las accesiones de *B. orellana*, en el que la accesión N4 tiene mayor contenido de bixina.

Las pruebas de Tetra Primer ARMS-PCR realizadas en los SNPs I (ID 4591), II (ID 7327), IV (ID 5465) y XV (ID 2875) mostraron amplificaciones similares para las accesiones N4 y N5. Por el contrario, la accesión P12 tuvo un patrón diferente a las accesiones N4 y N5.

Estos SNPs se relacionan con el color de la flor y la dehiscencia del fruto, puesto que, las accesiones N4 y N5 presentan flores rosas y fruto dehiscente; y la accesión P12 tiene flores blancas y fruto indehiscente.

5.2 PERSPECTIVAS

Conforme a la base de datos de SNPs que se pudo extraer de los transcriptomas, se pretende identificar los genes en donde se encuentran los SNPs seleccionados y analizarlos. Además, es posible hacer una matriz que señale el número de SNPs ubicados en los genes expresados en los transcriptomas, haciendo énfasis en los genes propios de la ruta de la bixina, y correlacionarlos al contenido de este.

En este trabajo se logró extraer una cantidad considerable de SNPs usando transcriptomas, por tanto, al analizar genomas completos el número de SNPs aumenta, así también se verá afectada la relación Ts/Tv y la tasa de variantes. Con esta información se podría continuar con estudios más profundos de SNPs comparando genomas de *B. orellana*, de un número de individuos mayor para poder tener un sustento estadístico y así conseguir SNPs estrechamente relacionados a características morfológicas de interés, no solo de bixina sino también de compuestos como geranil-geraniol, escualeno, b-sitosterol u otros que sean beneficiosos para la salud o la agricultura.

Con la información que se tiene de la técnica tetra-primer ARMS-PCR y los cebadores ya diseñados se procurará probarlos en más accesiones para su validación, una vez que se confirme su eficacia se pueden diseñar con certeza cebadores para genotipar SNPs extraídos del genoma completo de *B. orellana* de mayor confianza relacionados al compuesto deseado y en un futuro llevar a cabo un proceso de mejoramiento genético o la siembra de variedades para obtener plantaciones de achiote más capaces de producir compuestos esperados.

BIBLIOGRAFÍA

- Agner, A. R., Barbisan, L. F., Scolastici, C., y D. M. F. Salvadori (2004). Absence of carcinogenic and anticarcinogenic effects of annatto in the rat liver medium-term assay. *Food and Chemical Toxicology*, 42(10), 1687-1693.
- Arce Portuguez, J. 1984. Caracterización de 81 plantas de achiote (*Bixa orellana* L.) de la colección del CATIE procedentes de Honduras y Guatemala y propagación vegetativa por estacas. Tesis M.Sc. Turrialba, Costa Rica, CATIE. pp. 149.
- Avalos J. M. Caracterización morfológica de dos variedades de achiote (*Bixa orellana*) en el municipio de Comalcalco, Tabasco. Tesis como requisito parcial para obtener el grado de Magister en Sistemas Sustentable de Producción en el Trópico. Colegio de Postgraduados. México. pp. 50. 2013.
- Ayobamidele, A. (2018). Elevated carotenoids in staple crops: The biosynthesis, challenges and measures for target delivery. *Journal of Genetic Engineering and Biotechnology*. 16, 553–562.
- Azofeifa-Delgado, A. (2006). Uso de marcadores moleculares en plantas; aplicaciones en frutales del trópico. *Agronomía mesoamericana*. 17(2), 221-242.
- Beltran, J.C.M., Stange, C. (2016). Apocarotenoids: A New Carotenoid-Derived Pathway. In: Stange, C. (eds) Carotenoids in Nature. *Subcellular Biochemistry*, 79.
- Bohry, D., Ramos, H.C.C., dos Santos, P.H.D., Bastos, M. S., Santana, F. A., Vimercati, A. A. y Gonzaga, M. (2021). Discovery of SNPs and InDels in papaya genotypes and its potential for marker assisted selection of fruit quality traits. *Sci Rep* 11, 292.
- Bolger, A. M., Lohse, M. y Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120.
- Bouvier, F., Dogbo, O. y Camara, B. (2003). Biosynthesis of the Food and Cosmetic Plant Pigment Bixin (Annatto). *Science*, 300(5628), 2089–2091.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*. 234(2): 177-186.
- Brookes, A.J. The essence of SNPs. *Gene* 1999, 234, 177–186.

- Brumfield, R.T., Beerli, P., Nickerson, D. A. Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *TRENDS in Ecology and Evolution*. 18 (5), 240-256.
- Cárdenas-Conejo Y, Narváez-Zapata JA, Carballo-Uicab VM, Aguilar-Espinosa M, Us-Camas R, Escobar-Turriza P, Comai L and Rivera-Madrid R. (2023). Gene expression profile during seed development of *Bixa orellana* accessions varying in bixin pigment. *Front. Plant Sci*. 14, 1066509.
- Cárdenas-Conejo, Y., Carballo-Uicab, V., Lieberman, M., Aguilar-Espinosa, M., Comai, L. y Rivera-Madrid, R. (2015). De novo transcriptome sequencing in *Bixa orellana* to identify genes involved in methylerythritol phosphate, carotenoid and bixin biosynthesis. *BMC Genomics*, 16, 877.
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Current Opinion in Plant Biology*, 27, 207–216.
- Chiapparino, E., Lee, D., & Donini, P. (2004). Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. *Genome*, 47(2), 414–420.
- Cingolani, P. and Platts, A. and Coon, M. and Nguyen, T. and Wang, L. and Land, S.J. and Lu, X. and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 6(2), 80-92.
- Coelho, D., da Silva, A., Silva, J., Rodrigues, S. A., Diniz, A., Guerra, R. (2022). *Bixa orellana* L. (Achiote, Annatto) as an antimicrobial agent: A scoping review of its efficiency and technological prospecting. *Journal of Ethnopharmacology*. 287, 114961.
- Coelho, D. da Silva, A., Silva, J., Rodrigues, S. A., Diniz, A. y Guerra, R. (2022). *Bixa orellana* L. (Achiote, Annatto) as an antimicrobial agent: A scoping review of its efficiency and technological prospecting. *Journal of Ethnopharmacology*. 287, 114961.
- Cosgun, E., Oh, M. (2020). Exploring the Consistency of the Quality Scores with Machine Learning for Next-Generation Sequencing Experiments. *BioMed Research International*, 2020, 8531502.
- Cubero, I. J. (2002). Marcadores y Mapas Genéticos, Introducción a la Mejora Genética Vegetal. 2a Edición, Mundi-Prensa, Madrid, 3, 67-100.

- Duangjit J, Bohanec B, Chan AP, Town CD, Havey MJ. (2013). Transcriptome sequencing to produce SNP-based genetic maps of onion. *Theor Appl Genet.* 126(8), 2093-101.
- Duarte, J., Rivière, N., Baranger, A. et al. (2014). Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics.* 15, 126.
- Duque Y Duque, E. R., Aguirre Saltos, M. B., & Tamayo Domínguez, A. C. (2022). Caracterización Fenotípica, Genotípica y ensayos de autopolinización en 18 accesiones de Achiote (*Bixa orellana* L.) en Costa Rica. *Agronomía Costarricense,* 46(2).
- Echevarría-Machado, I., Sánchez-Cach, L.A., Hernández-Zepeda, C., Rivera-Madris, R., Moreno-Valenzuela, O. A. (2005). A simple and efficient method for isolation of DNA in high mucilaginous plant tissues. *Mol Biotechnol.* 31, 129–135.
- Ellison, S. L. (2016). Carotenoids: Physiology. *Encyclopedia of Food and Health.* 670–675.
- Escobar-Turriza, P. J. 2017. Identificación de marcadores moleculares para el mejoramiento genético del achiote (*Bixa orellana* L.). Tesis de maestría. Centro de Investigación Científica de Yucatán.
- Frazer, P. D. y Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. *Progress in Lipid Research.* 43, 228–265.
- Frusciante, S., Demurtas, O. C., Sulli, M., Mini, P., Aprea, G., Diretto, G., Karcher, D., Bock, R. y Giuliano, G. (2022). Heterologous expression of *Bixa orellana* cleavage dioxygenase 4–3 drives crocin but not bixin biosynthesis, *Plant Physiology,* 188(3), 1469–1482.
- Garcés-Claver, A., Fellman, S.M., Gil-Ortega, R. (2007). Identification, validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in *Capsicum* spp.. *Theor Appl Genet.* 115, 907–916.
- Garcés-Claver, A., Fellman, S.M., Gil-Ortega, R. et al. (2007). Identification, validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in *Capsicum* spp.. *Theor Appl Genet* 115, 907–916.

- Geraldes, A., DiFazio, S. P., Slavov, G. T., Ranjan, P., Muchero, W., Hannemann, J., ... Tuskan, G. A. (2013). A 34K SNP genotyping array for *Populus trichocarpa*: Design, application to the study of natural populations and transferability to other *Populus* species. *Molecular Ecology Resources*, 13(2), 306–323.
- Isaza, I. I., Jaramillo, T. J., Posada, L. F. y Palacio, G. L. (2019). Evaluación del proceso de extracción alcalina de bixina de semillas de achiote (*Bixa orellana*). Proyecto de Grado. Universidad EAFIT. Ingeniería de Procesos.
- Javaid, N., Ramzan, M., Khan, I.A. et al. (2022). The chloroplast genome of *Farsetia hamiltonii* Royle, phylogenetic analysis, and comparative study with other members of Clade C of Brassicaceae. *BMC Plant Biol*, 22, 384.
- Kim, T.-H.; Yoon, Y.-M.; Park, J.-C.; Park, J.-H.; Kim, K.-H.; Kim, Y.-K.; Son, J.-H.; Park, T.-I. (2022). De Novo Transcriptome Assembly and SNP Discovery for the Development of dCAPS Markers in Oat. *Agronomy*. 12, 184.
- Koc, I., Filiz, E. y Tombuloglu, H. (2015). Comparative analysis of plant lycopene cyclases. *Computational Biology and Chemistry*, 58, 81–92.
- Langmead B y Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359.
- Li, H. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- Lohr, M. (2009). Carotenoids. *The Chlamydomonas Sourcebook*. 799–817.
- López, S. E., Caicedo, M., Gil, A., López, A., & Pazos, A. E. (2018). Morfometría de fruto y semilla de *Bixa orellana* L. “achiote”. *SCIÉND*O, 21(2), 213-216.
- López-Medina, S. E., Caicedo, M., Gil, A., López, A., & Pazos, A. E. (2018). Morfometría de fruto y semilla de *Bixa orellana* L. “achiote”. *SCIÉND*O. 21(2), 213-216.
- Martín-Rodríguez, E. (2010). Uso de genes involucrados en la síntesis de carotenoides como marcadores moleculares de pigmentos en achiote (*Bixa orellana* L.). Tesis de maestría. Centro de Investigación Científica de Yucatán. Yucatán, México.
- Medrano, R.F.V., de Oliveira, C.A. (2014). Guidelines for the Tetra-Primer ARMS-PCR Technique Development. *Mol Biotechnol*, 56, 599–608.

- Misawa, N. (2010). 1.20 – Carotenoids. *Comprehensive Natural Products II*, 1, 733-753.
- Moreno JC, Pizarro L, Fuentes P, Handford M, Cifuentes V, Stange C. (2013). Levels of lycopene β -cyclase 1 modulate carotenoid gene expression and accumulation in *Daucus carota*. *PLoS One*. 8(3), 58144.
- Moreno, J.C., Mi, J., Alagoz, Y. y Babili, S. A. (2020). Plant apocarotenoids: from retrograde signaling to interspecific communication. *The Plant Journal*, 105(2), 351-375.
- Morgil H, Can Gercek Y and Tulum I (2020) Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. *The Recent Topics in Genetic Polymorphisms. IntechOpen*.
- Morin, P. A., Luikart, G. y Wayne R. K. (2004). SNPs in ecology, evolution and conservation. *TRENDS in Ecology and Evolution*. 19(4), 208-216.
- Morton J (1960) Can annatto (*Bixa orellana* L.), an old source of food color, meet new needs for safe dye?. *Proceedings of the Florida State Horticultural Society*. 73, 301–309.
- Nosková, A., Li, C., Wang, X., Leonard, A., Pausch, H., Kadri, N. K. (2023). Exploiting public databases of genomic variation to quantify evolutionary constraint on the branch point sequence in 30 plant and animal species. *Nucleic Acids Research*. 51(22),12069–12075.
- Orwa C, A Mutua, Kindt R, Jamnadass R, S Anthony. (2009). Agroforestry Database:a tree reference and selection guide version 4.0.
- Park, H., Kim, S., Nie, H., Kim, J., Lee, J., & Kim, S. (2020). Molecular identification of sweet potato accessions using ARMS-PCR based on SNPs. *Journal of Plant Biotechnology*, 47(2), 124–130. <https://doi.org/10.5010/JPB.2020.47.2.124>
- Pech-Hoil, R. (2020). Caracterización genética y molecular de una colección de achiote (*Bixa orellana* L.) en Yucatán. Centro de Investigación Científica de Yucatán. México.
- Pinzón L. L. L., C. H. Avendaño A., J. O. Mijangos C., G. del C. Godoy H., D. E. Cituk C., P. J. Correa N. & E. Campos R. (2008). Situación del recurso achiote en México.
- Quintos-Cortes, M. G., Fernández-Lozada, O. F., Petrolí, C.D., Barrozo-Toledo, F. H., Romero-Cortes, T., Cuervo-Parra, J.A. y Peralta-Gil, M. (2021). Desarrollo de

- marcadores moleculares de alto rendimiento y análisis de datos genotípicos en plataformas de avanzada. *Exploratoris: Revista de la Realidad Global*, 10, (1).
- Quiñones, B. X y Yunda-Romero M. C. (2014). El achiote *Bixa orellana* L. como posible alternativa productiva para el Departamento del Meta. *Rev Sist Prod Agroecol.* 5, 1.
- Raddatz-Mota, D., Pérez-Flores, LJ, Carrari, F. et al. (2017). Achiote (*Bixa orellana* L.): a natural source of pigment and vitamin E. *Journal of Food Science and Technology.* 54, 1729–1741
- Rentaría Alcántara, M. (2007). Capítulo 18. Breve revisión de los marcadores moleculares. En L. E. Eguiarte, V. Souza, & X. Aguirre, *Ecología Molecular*, México. pp. 541-562.
- Riju, A. y Arunachalam, V. (2010). Electronic Sorting of SNP/Indel Sites in Expressed Sequence Tag Libraries of Cocoa (*Theobroma cacao* L.). *Genes, Genomes and Genomics.*
- Rivera-Madrid R, Aguilar-Espinosa M, Ca´rdenas-Conejo Y, Garza-Caligaris LE (2016) Carotenoid derivatives in achiote (*Bixa orellana*) seeds: synthesis and health promoting properties. *FrontPlant Sci*, 7, 1406.
- Rivera-Madrid, R. (2021). La cadena de valor del achiote (*Bixa orellana*) detrás del avance científico y tecnológico. *Desde el Herbario CICY.* 13: 222–226.
- Sierra-Ulín, D. L. (2021). Análisis de la posible participación de los genes ALDH3H1-1 y SABATH4 en la biosíntesis de bixina en *Bixa orellana* L. Tesis de maestría. Centro de Investigación Científica de Yucatán.
- Silva-Junior, O. B., Faria, D. A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. *New Phytol.* 206, 1527–1540.
- Singh, M., Kumar, S. (2023). Effect of single nucleotide polymorphisms on the structure of long noncoding RNAs and their interaction with RNA binding proteins. *Biosystems*, 233, 105021.

- Siva, R., Doss, F. P., Kundu, K., Satyanarayana, V. S. V., & Kumar, V. (2010). Molecular characterization of bixin—An important industrial product. *Industrial Crops and Products*, 32(1), 48–53.
- Smith J. (2006). Annatto extracts-chemical and technical assessment. Chem Tech Assess Manual. 1–21.
- SNICS (2017). Achiote (*Bixa orellana* L.): Generalidades de la Red Achiote. <https://www.gob.mx/snics/acciones-y-programas/achiote-bixa-orellana-l>
- Tamayo-García, R., Narváez-Zapata, J.A., Ku-González, A. et al. (2022). Gene expression analysis during the fruit development in dehiscent and indehiscent *Bixa orellana* L. accessions. *Physiol Mol Biol Plants* 28, 709–718.
- Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, Wu S, Wang Y. (2023). SRplot: A free online platform for data visualization and graphing. *PLoS One*. 18(11), 0294236.
- Teixeira da Silva, J.A., Zeng, S., Godoy-Hernández, G., Rivera-Madrid, R., Dobránszki, J. (2019). *Bixa orellana* L. (achiote) tissue culture: a review. *In Vitro Cellular & Developmental Biology – Plant*. 55, 231–241.
- Trujillo-Hernández, J. A. (2013). Caracterización de la variación alélica en genes involucrados en la biosíntesis de carotenoides en *Bixa orellana* L. Tesis de Maestría. Centro de Investigación Científica de Yucatán.
- Trujillo-Hernández, J.A., Cárdenas-Conejo, Y., Turriza, P.E., Aguilar-Espinosa, M., Carballo-Uicab, V., Garza-Caligaris, L.E., Comai, L. y Rivera-Madrid, R. (2016). Functional polymorphism in lycopene beta-cyclase gene as a molecular marker to predict bixin production in *Bixa orellana* L. (achiote). *Mol Breeding*. 36, 135.
- Us-Camas R, Aguilar-Espinosa M, Rodríguez-Campos J, Vallejo-Cardona AA, Carballo-Uicab VM, Serrano-Posada H and Rivera-Madrid R (2022). Identifying *Bixa orellana* L. New Carotenoid Cleavage Dioxygenases 1 and 4 Potentially Involved in Bixin Biosynthesis. *Front. Plant Sci*. 13, 829089.
- Vilar de A, Vilar MS, de Lima e Moura TF, Raffin FN, de Oliveira MR, Franco CF, de Athayde-Filho PF, Diniz Mde F, Barbosa-Filho JM. (2014). Traditional uses, chemical

- constituents, and biological activities of *Bixa orellana* L.: a review. *The Scientific World Journal*. 2014, 857292.
- Walter, M. H. y Strack, D. (2011). Carotenoids and their cleavage products: Biosynthesis and functions. *Natural Product Reports*. 28, 663.
- Wang J, Raskin L, Samuels DC, Shyr Y, Guo Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*. 31(3),318-23.
- Yang, Z., Lo, Y.T., Quan, Z. et al. (2023). Application of a modified tetra-primer ARMS–PCR assay for rapid *Panax* species identity authentication in ginseng products. *Sci Rep* 13, 14396.
- Ye, S., Dhillon, S., Ke, X., Collins, A. R., & Day, I. N. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*. 29, 88.
- Valdez-Ojeda, R., Quiros, C. F., Aguilar-Espinosa, M. de L., & Rivera-Madrid, R. (2010). Outcrossing rates in annatto determined by sequence-related amplified polymorphism. *Agronomy Journal*, 102(5), 1340–1345. <https://doi.org/10.2134/agronj2009.0510>