



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

Estudio de la expresión de genes relacionados con el metabolismo de capsaicinoides en frutos de plantas de chile habanero (*Capsicum chinense* Jacq.) cultivadas en diferentes tipos de suelo

Tesis que presenta

Eduardo José Burgos Valencia

En opción al título de

DOCTORADO EN CIENCIAS

(Ciencias Biológicas: BIOQUÍMICA Y BIOLOGÍA MOLECULAR)

Mérida, Yucatán, México

2022


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 de **Eduardo José Burgos Valencia** titulado “**Estudio de la expresión de genes relacionados con el metabolismo de capsaicinoides en frutos de plantas de chile habanero (*Capsicum chinense* Jacq) cultivadas en diferentes tipos de suelo**”, fue realizado en los laboratorios de la **Unidad de Bioquímica y Biología Molecular de Plantas del Centro de Investigación Científica de Yucatán, A.C.** bajo la dirección del **Dr. Manuel Martínez Estévez** y la co-dirección del **Dr. José Alberto Narváez Zapata**, dentro de la opción de Bioquímica y Biología Molecular, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente,

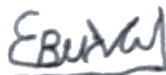


Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 4 de noviembre de 2022

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos, los 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 de 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 en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.



M. en C. Eduardo José Burgos Valencia

Este trabajo se llevó a cabo en la Unidad de Bioquímica y Biología Molecular de Plantas del Centro de Investigación Científica de Yucatán, A.C., y forma parte del proyecto titulado **ANÁLISIS DE LOS CAMBIOS METABOLÓMICOS DURANTE EL DESARROLLO DEL FRUTO DE *CAPSICUM CHINENSE* JACQ. CULTIVADO EN DIFERENTES TIPOS DE SUELO**, con el número 000000000257588, a nombre de la doctora Ingrid Mayanin Rodríguez Buenfil, y bajo la dirección del Dr. Manuel Martínez Estévez y el Dr. José Alberto Narváez Zapata.

AGRADECIMIENTOS

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada (Registro de becario 711358) a lo largo de mis estudios de doctorado.

Agradezco al Centro de Investigación Científica de Yucatán (CICY) por haberme dado la oportunidad de estudiar el doctorado y de trabajar en sus instalaciones y con sus equipos, y el personal de apoyo, así como por la formación académica y el aprendizaje adquirido durante la realización de mi tesis.

A mis directores de tesis el doctor Manuel Martínez Estévez y el doctor José Alberto Narváez Zapata por haberme aceptado como su estudiante, así como por su enseñanza, apoyo, guía y conocimientos brindados que permitieron mi formación académica.

Agradezco al Centro de Biotecnología Genómica (CBG) del Instituto Politécnico Nacional (IPN) por darme la oportunidad de realizar una estancia en sus instalaciones y participar en su curso de Bioinformática, y a la Escuela Superior de Medicina del IPN por permitirme participar en su curso de Modelado Molecular, todo lo cual me brindó conocimientos que me ayudaron a finalizar el proyecto.

A mis asesores de tesis y miembros de mi comité predoctoral, la doctora Ileana Echevarría Machado y la doctora Ingrid Mayanin Rodríguez Buenfil, por su apoyo, conocimientos, seguimiento e interés en la realización del proyecto y mi formación académica. Asimismo, a los otros miembros de mi comité predoctoral y revisores, el doctor Luis Carlos Rodríguez Zapata, el doctor Felipe Sánchez Teyer y la doctora Georgina Estrada Tapia, por sus conocimientos y valiosas aportaciones durante la realización de este trabajo. También al doctor Manuel Octavio Ramírez Sucre por sus valiosas aportaciones y correcciones durante la revisión de este trabajo.

A la maestra Fátima Medina Lara, el maestro Ramón Souza Perera y la maestra Miriam Monforte González por su apoyo técnico, paciencia, disposición, interés y conocimientos transmitidos, todo lo cual permitió el avance y finalización de esta tesis.

Al doctor Gustavo Ortega-Lule por su apoyo y conocimientos en el modelamiento de las proteínas que se realizó en este trabajo, que aportó de manera valioso a alcanzar los objetivos de tesis.

A todos mis compañeros del CICY, que de alguna manera contribuyeron a la realización de este trabajo. Por todos los buenos momentos que hemos pasado y su convivencia, que me ha ayudado en el crecimiento académico y personal. En particular, a mi compañero el maestro Federico García Laynes, por su apoyo en las extracciones de RNA. Mis más sinceros agradecimientos a todos.

A todos mis amigos, gracias.

DEDICATORIAS

A mis padres. Por su apoyo y amor en los momentos difíciles para seguir adelante, y sus enseñanzas y educación que me han formado de manera personal durante mis 31 años de vida.

A mis familiares. Por todo el amor que me han dado.

A Dios, por darme fortaleza para la realización de esta tesis continuar adelante a lo largo de mi vida.

Al doctor Manuel Martínez Estévez y el doctor José Alberto Narvárez Zapata, por sus conocimientos, apoyo, guía y dedicación brindados durante la realización de esta tesis.

A la doctora Ileana Echevarría Machado por su interés y dedicación en el seguimiento de mi formación académica y para la finalización de este trabajo.

A todos mis amigos y a mis compañeros de estudio que me han apoyado durante mi vida y durante la realización de este trabajo. Por los buenos momentos que hemos pasado, y su convivencia en las buenas y las malas, que me han ayudado a crecer como profesional y como persona.

PRODUCTOS ASOCIADOS

Artículo de revisión titulado “Gene expresión related to the capsaicinoid biosynthesis in the *Capsicum* genus: Molecular and transcriptional studies”, publicado en el Brazilian Journal of Botany (<https://doi.org/10.1007/s40415-019-00575-6>).

Resumen y cartel titulado “Caracterización de la floración y frutos de plantas de chile habanero crecidas en diferentes tipos de suelo de Yucatán” en el Segundo Congreso de Fisiología Vegetal del 6 al 8 de noviembre de 2019.

Artículo de divulgación titulado “El chile habanero y la importancia de los diferentes tipos de suelo de Yucatán en su cultivo”, publicado en el Hasnup del CICY de julio de 2022.

ÍNDICE

ÍNDICE DE FIGURAS	v
ÍNDICE DE TABLAS	vii
LISTADO DE ABREVIATURAS	viii
INTRODUCCIÓN	1
CAPÍTULO I	3
ANTECEDENTES	3
1.1. Género <i>Capsicum</i>. Generalidades	3
1.2. El chile habanero (<i>Capsicum chinense</i> Jacq.)	3
1.3. Factores que influyen sobre las características fenotípicas del chile habanero ...	4
1.4. El género <i>Capsicum</i> y los capsaicinoides	6
1.5. Factores que influyen en el contenido de capsaicinoides	9
1.6. Los suelos cultivables en Yucatán	12
1.7. Expresión de genes y transcriptómica de la ruta d biosíntesis de capsaicinoides	14
PREGUNTA DE INVESTIGACIÓN	18
HIPÓTESIS	18
JUSTIFICACIÓN	19
OBJETIVO GENERAL	19
OBJETIVOS ESPECÍFICOS	20
ESTRATEGIA EXPERIMENTAL	20
CAPÍTULO II. Gene expression related to the capsaicinoid biosynthesis in the <i>Capsicum</i> genans: a molecular and transcriptomic approaches	22
ABSTRACT	22
2.1. INTRODUCTION	22

ÍNDICE

2.2. GENE EXPRESSION AND TRANSCRIPTOME ANALYSIS.....	24
2.3. STUDIES OF GENE EXPRESSION AND TRANSCRIPTOMICS RELATED TO CAPSAICINOIDS IN THE GENUS <i>Capsicum</i>	25
CAPÍTULO III. Expresión diferencial de genes relacionados con el desarrollo y síntesis de capsaicinoides en frutos de plantas de chile habanero (<i>Capsicum chinense</i> Jacq.) crecidas en diferentes tipos de suelo	39
RESUMEN	39
3.1. INTRODUCCIÓN	40
3.2. METODOLOGÍA.....	42
3.2.1. Condiciones de crecimiento de las plantas	42
3.2.2. Etiquetado de flores, colecta y caracterización de frutos	43
3.2.3. Determinación de capsaicinoides.....	44
3.2.4. Análisis estadístico	45
3.2.5. Extracción de RNA	45
3.2.6. Secuenciación del transcriptoma de frutos de chile habanero.....	46
3.2.7. Filtrado de <i>reads</i> y ensamblado.....	46
3.2.8. Predicción de ORFs, anotación funcional y estimación de la abundancia de los unigenes	47
3.2.9. Análisis de expresión diferencial y de rutas metabólicas enriquecidas de los unigenes	47
3.2.10. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides	48
3.3. RESULTADOS	49
3.3.1. Producción de flores y frutos y caracterización de frutos en plantas de chile habanero crecidas en dos tipos de suelo	49
3.3.2. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en dos tipos de suelo	53
3.3.3. Análisis general del transcriptoma y de rutas metabólicas enriquecidas en frutos de plantas de chile habanero crecidas en dos tipos de suelo	55

3.3.4. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides	62
3.4. DISCUSIÓN	69
3.4.1. Producción de flores y frutos y caracterización de frutos en plantas de chile habanero crecidas en dos tipos de suelo	69
3.4.2. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en dos tipos de suelo	71
3.4.3. Análisis general del transcriptoma y de rutas metabólicas enriquecidas en frutos de plantas de chile habanero crecidas en dos tipos de suelo	73
3.4.4. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides	76
3.5. CONCLUSIONES	80
CAPÍTULO IV. Haplotype analysis, regulatory elements and docking simulation of structural models of different <i>AT3</i> copies in the genus <i>Capsicum</i>	82
ABSTRACT	82
4.1. INTRODUCTION	83
4.2. MATERIAL AND METHODS	84
4.2.1. <i>AT3</i> orthologues identification and characterization	84
4.2.2. Genomic regulatory elements	85
4.2.3. Modeling the <i>AT3</i> protein	86
4.2.4. Molecular docking of the <i>AT3</i> protein	86
4.3. RESULTS	87
4.3.1. <i>AT3</i> gene analysis	87
4.3.2. Haplotype network analysis	89
4.3.3. Regulatory elements analysis	91
4.3.4. Modeling the <i>AT3</i> protein	94
4.3.5. Docking simulation with probable substrates	98
4.4. DISCUSSION	100

ÍNDICE

4.4.1. AT3 phylogenetic and haplotype analysis	100
4.4.2. Regulatory elements	101
4.4.3. AT3 protein model	103
4.4.4. Docking analysis	104
4.5. CONCLUSION	106
CAPÍTULO V.....	107
5.1. DISCUSIÓN GENERAL	107
5.2. CONCLUSIONES GENERALES.....	114
5.3. PERSPECTIVAS.....	115
BIBLIOGRAFÍA.....	116
ANEXOS	144
TABLAS SUPLEMENTARIAS.....	144
FIGURAS SUPLEMENTARIAS.....	171

ÍNDICE DE FIGURAS

Figura 1.1. Estructura química de los capsaicinoides8

Figura 1.2. Ruta de biosíntesis de capsaicinoides 10

Figura 1.3. Esquema de la estrategia experimental.....21

Figure 2.1. Capsaicinoid biosynthesis model.....31

Figura 3.1. Disposición de las plantas de chile habanero en el invernadero43

Figura 3.2. Producción de flores y frutos en las plantas de chile habanero crecidas en suelos negro y rojo52

Figura 3.3. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en suelos negro y rojo.....55

Figura 3.4. Análisis de las categorías funcionales principales GO de los *contigs* del transcriptoma de frutos58

Figura 3.5. Análisis global de la expresión diferencial de genes y de las rutas metabólicas enriquecidas del transcriptoma de frutos en suelo negro y rojo62

Figura 3.6. Análisis de la expresión de genes de las rutas metabólicas enriquecidas y de la síntesis de capsaicinoides del transcriptoma de frutos en suelo negro y rojo69

Figure 4.1. Phylogenetic tree of the AT3 sequences in *Capsicum* and homologues89

Figure 4.2. Haplotype map of the AT3 sequences in *Capsicum*.....91

Figure 4.3. Structure models and surfaces of AT3-D1 and AT3-D2.98

Figure 4.4. Docking results of AT3 models with the CS substrates. 100

Figura 5.1. Resumen general..... 113

Figura Suplementaria 3.1. Corrida de muestras de capsaicina en frutos de chile habanero por HP-TLC 171

ÍNDICE DE FIGURAS

Figura Suplementaria 3.2. Calidad por base en los <i>reads</i> en el transcriptoma de frutos de 25 DPA de plantas de chile habanero	172
Figura Suplementaria 3.3. Distribución de la calidad en los <i>reads</i> del transcriptoma de frutos de 25 DPA de plantas de chile habanero	174
Figura Suplementaria 3.4. Distribución de la longitud de los <i>reads</i> del transcriptoma de frutos de 25 DPA de plantas de chile habanero	176
Figura Suplementaria 3.5. Análisis de enriquecimiento de genes por términos KEEG en la comparación de la expresión de los genes del transcriptoma en suelo rojo contra suelo negro	177
Figura Suplementaria 3.6. Mapas de calor de expresión de genes de frutos de chile habanero en suelos negro y rojo.....	179
Figura Suplementaria 3.7. Análisis filogenético de los ortólogos de <i>AT3</i> en el transcriptoma de frutos de chile habanero	180
Figura Suplementaria 4.1. Alineamiento múltiple de secuencias de <i>AT3</i> y homólogos.	181
Figura Suplementaria 4.2. Evaluación de <i>AT3-D1</i> y <i>AT3-D2</i> por comparación con sus templates..	182
Figura Suplementaria 4.3. Resultados de docking de los modelos de <i>AT3</i> con los sustratos de DAT.....	183

ÍNDICE DE TABLAS

Tabla 1.1 Clasificación taxonómica del chile habanero.....	4
Table 2.1. Studies related to genes and molecular findings of capsaicinoid synthesis pathway	36
Table 2.2. Transcriptomic studies related with capsaicinoid synthesis pathway	37
Tabla 3.1. Estadísticas de los <i>reads</i> filtrados del transcriptoma de frutos de 25 DPA.....	55
Tabla 3.2. Estadísticas del ensamble del transcriptoma de frutos de 25 DPA.....	56
Tabla 3.3. <i>Contigs</i> usados para el análisis de expresiónl	63
Table 4.1. Regulatory elements at 2000 pb upstream of the start of <i>AT3-D1</i> and <i>AT3-D2</i>	92
Table 4.2. General confidence values of AT3-D1 and AT3-D2 models and their docking interaction with different substrates	96
Tabla suplementaria 3.1. Información de genes diferencialmente expresados en frutos de 25 DPA de plantas de chile habanero crecidas en suelo rojo o negro	144
Tabla Suplementaria 3.2. Anotación funcional de genes seleccionados para el análisis de expresión y datos de expresión de los genes en TPM	148
Tabla Suplementaria 3.3. Capsaicinoides en frutos de chile habanero en SHU	153
Tabla Suplementaria 3.4. Frutos totales de 45 DPA de chile habanero en diferentes estados de madurez.....	153
Tabla Suplementaria 4.1. Secuencias de <i>AT3</i> y homologos en el genero <i>Capsicum</i> obtenidas de la base de datos del NCBI	153
Tabla suplementaria 4.2. Detalles de las regiones reguladoras a 2000 pb río arriba del inicio de <i>AT3-D1</i> and <i>AT3-D2</i>	159
Tabla Suplementaria 4.3. Análogos estructurales hallados por I-TASSER con la similitud estructural más cercana con AT3-D1 y AT3-D2	169
Tabla Suplementaria 4.4. Propiedades de bolsillo del canal de solvente de AT3	170

LISTADO DE ABREVIATURAS

LISTADO DE ABREVIATURAS

Abreviatura	Significado
6PGD	6-fosfogluconato deshidrogenasa
AK	Aspartato cinasa
ANOVA	Análisis de varianza
ARG1	Arginasa 1
AT3	Aciltransferasa, gen de la capsaicinoide sintasa
AT3-D1	Haplotipo del grupo 1 de AT3
AT3-D2	Haplotipo del grupo 2 de AT3
BBH	Bi-directional best hit
BP	Proceso biológico
CC	Componente celular
cDNA	DNA complementario
CDS	Secuencia codificante de un gen
CS	Capsaicinoide sintasa
DAF	Días después de la floración
DPA	Días post-antesis
DGE	Análisis diferencial de expresión de genes
DPT	Días post-trasplante
DXS	1-Desoxi-xilulosa-5-fosfato sintasa
DXS2	1-Desoxi-xilulosa-5-fosfato sintasa 2
EST	Expressed sequece tag
FAT	Acil-ACP tioesterasa
FPKM	Fragmentos por kilobase de transcrito por millones de <i>reads</i> mapeados
G6PDH	Glucosa-6-fosfato deshidrogenasa
GABAT1	Gamma-aminobutirato transaminasa 1
GABAT3	Gamma-aminobutirato transaminasa 3
GluTR	Glutamil-tRNA reductasa
GO	Gene Ontology
GOGAT	Glutamato sintasa
GS	Glutamina sintetasa
GSA	Glutamato-1-semialdehído 2,1-aminomutasa
GSEA	Análisis de enriquecimiento de sets de genes
HPTLC	Cromatografía de capa fina de alta resolución
KARI	Cetol-ácido reductoisomerasa
KAS	Cetoacil-ACP sintasa
KCS	Cetoacil-CoA sintasa
KEGG	Kyoto Encyclopedia of Genes and Genomes
KR	Acil-ACP reductasa
LFC	Logaritmo base 2 de la relación de expresión de genes entre suelos
LHCI	Complejo de recolección de luz del fotosistema I
LHCII	Complejo de recolección de luz del fotosistema II
LHCA2	Proteína 2 de unión a clorofila del complejo LHCI
LHCB3	Proteína 3 de unión a clorofila del complejo LHCII
LHY	Hipocótilo alargado tardío
MF	Función molecular

LISTADO DE ABREVIATURAS

NSG	Tecnologías de secuenciación de nueva generación
ORF	Marco de lectura abierto
OTC	Ornitina transcarbamilasa
pAMT	Aminotransferasa putativa (vanillilamina)
PCR	Reacción en cadena de la polimerasa
PDS	15-cis fitoeno desaturasa
POD	Peroxidasas
PRR37	Regulador de pseudorespuesta 7
PS	Peso seco (fruto)
PSI	Fotosistema I
PSII	Fotosistema II
PSY2	15-cis fitoeno sintasa 2
PUN1	Gen/locus de la capsaicinoide sintasa
QLT	Loci de rasgo cuantitativo
RPN2	Subunidad N2 regulatoria del proteasoma 26S
RPT4	Subunidad T4 regulatoria del proteasoma 26S
RuBisCO	Ribulosa-1,5-bifosfato carboxilasa/oxigenasa
SHU	Unidades de Scoville
SN	Suelo negro
SR	Suelo rojo
TPM	Transcritos por millón
UGT	UDP-glucosiltransferasa

RESUMEN

RESUMEN

El chile habanero (*Capsicum chinense* Jacq.) es una planta de importancia económica en la Península de Yucatán. Parte importante de la calidad del chile habanero es su alto contenido de capsaicinoides que les dan el picor a los frutos. Estos metabolitos son de importancia comercial por la sensación de pungencia y sus propiedades benéficas para la salud. Diferentes factores ambientales contribuyen en la calidad de los frutos y en su acumulación de capsaicinoides, entre los que destaca el tipo de suelo donde crece el chile habanero. A pesar de su importancia en el cultivo de chile, su efecto en el metabolismo de capsaicinoides apenas es entendido. El presente trabajo tuvo como objetivo comparar el efecto de suelos con características contrastantes, suelo negro (SN) y rojo (SR), sobre el contenido de capsaicinoides y la expresión de genes relacionados con el metabolismo de éstos por medio de transcriptómica. El SR fue favorable para la obtención de un mayor número de frutos más desarrollados y mayor expresión de genes relacionados con el desarrollo floral y la madurez de los frutos. En SN se vio favorecida la acumulación de capsaicinoides en frutos de 25 DPA, y expresión más alta de genes relacionados con la fotosíntesis, la clorofila, la ruta de las pentosas fosfato, el metabolismo de aminoácidos, terpenoides y con el proteasoma. Se encontró mayor expresión de genes relacionados con la síntesis de valina y leucina, la ruta de las pentosas fosfato y el proteasoma con posible relación con la acumulación de capsaicinoides. De entre los genes analizados, el gen *AT3* de la capsaicinoide sintasa (CS) ha sido reportado como regulatorio en la síntesis de capsaicinoides. Dos haplotipos de *AT3* con diferente nivel de expresión en el transcriptoma fueron analizados. El modelo proteico de *AT3* fue similar a la de las aciltransferasas de la familia BAHD. Se encontraron diferencias en la forma, tamaño y composición del canal de solvente entre estructuras de los diferentes modelos de *AT3*. Asimismo, se encontraron diferentes afinidades con sustratos de la CS: (E)-8-metilnon-6-enoil-CoA, 8-metilnonanoil-CoA y vanillilamina. El análisis de genómico entre haplotipos mostró diferencias entre elementos regulatorios relacionados principalmente con luz y estrés abiótico. Por lo que se sugiere que estos haplotipos se expresan diferencialmente ante estímulos abióticos similares. Estos resultados muestran que el tipo de suelo tiene un efecto sobre el tipo de procesos que se favorecen en los frutos de chile habanero, que incluye a los genes relacionados con el metabolismo de capsaicinoides. Además, la importancia en la función de éstos también debe profundizarse a nivel de estructura proteica y actividad.

Palabras clave: *Capsicum chinense*, suelo, capsaicinoides, expresión de genes, capsaicinoide sintasa

ABSTRACT

Habanero pepper (*Capsicum chinense* Jacq.) is a crop of economic importance in the Yucatan Peninsula. Part of the quality of the habanero pepper is its high capsaicinoid content, which give the spicy sensation of its fruits. These metabolites are commercially relevant due its pungency and its healthy properties. Environmental factors contribute to the quality of the habanero pepper fruits and in the capsaicinoid accumulation, among them the soil of cultivation have a central importance. Despite its relevance in the pepper cultivation, its effect in the metabolism of capsaicinoids its barely understood. The aim of this work was the comparison of the effect of two types of soil which contrasting features, black soil (BS) and red soil (RS), in the capsaicinoid accumulation and in the expression of the genes related with the capsaicinoid metabolism using a transcriptomic analysis. RS was favorable to obtain bigger fruits and higher expression of genes related with floral development and in the ripe of the fruits. BS was favorable for the capsaicinoid accumulation in fruits of 25 DAA, and higher expression related to photosynthesis, chlorophyll, pentose phosphate pathway, aminoacids and terpenoids metabolism, and the proteasome. Specifically, there was higher expression of genes related with valine and leucine synthesis, pentose phosphate pathway and of the proteasome with a possible relation with capsaicinoid accumulation. Among the analyzed genes, the *AT3* gene of the capsaicinoid synthase (CS) has been reported as regulatory in the capsaicinoid biosynthesis. Two haplotypes of *AT3* with different expression level founded in the transcriptome were analyzed. The protein model of *AT3* was similar to the acyltransferases of the BAHD superfamily. Differences were found in the shape, size and composition of the solvent channel between the different *AT3* structural models. Also, different interaction affinities with CS substrates were found: (E)-8-methylnon-6-enoyl-CoA, 8-methylnonanoyl-CoA and vanillylamine. Genomic analysis between haplotypes found different regulatory upstream regions mainly related with light and abiotic stress, which suggest that these haplotypes respond with differential expression by the same abiotic stimulus. The obtained results demonstrate that the soil type had an effect in the type of development process in the habanero pepper fruits, that includes the expression of genes related with the capsaicinoid metabolism. Besides, the importance of them must be analyzed in proteic structure and activity level.

Keywords: *Capsicum chinense*, soil, capsaicinoids, gene expression, capsaicinoid synthase

INTRODUCCIÓN

El chile habanero (*Capsicum chinense* Jacq.) es un cultivo tradicional en el sureste de México y, en particular, de la península de Yucatán. Este cultivo tiene una gran importancia económica y cultural y produce una gran variedad de metabolitos secundarios incluyendo a los capsaicinoides (Antonio *et al.*, 2018; Aranha *et al.*, 2017). Estos últimos compuestos son de gran interés ya que dan a los frutos de chile su picor característico.

Además, poseen propiedades bioactivas que los hacen de interés en industrias como la farmacéutica (Ghiasi *et al.*, 2019; Clark y Lee, 2016; Srinivasan, 2016). Por sus características de sabor, color, aroma, vida de anaquel y picor, el chile habanero producido en la península de Yucatán es considerado uno de los de mejor calidad y de los más picantes del mundo (Ruiz-Lau *et al.*, 2011; Canto-Flick *et al.*, 2008). Dadas su importancia económica y cultural, y por sus características, este cultivo obtuvo la denominación de origen como “Chile habanero de la Península de Yucatán”.

Las características particulares del chile habanero de Yucatán se deben en parte importante a las condiciones medioambientales en las que crecen las plantas. El genotipo, el ambiente y la interacción de estos factores influyen en el desarrollo del fruto y en su contenido químico, incluyendo a los capsaicinoides (Jeeatid *et al.*, 2018a). El suelo es uno de los factores ambientales más importantes a considerar en los cultivos.

Los suelos con características diferentes, como la cantidad de nutrientes, agua, salinidad, profundidad, partículas que los conforman, o poblaciones microbiológicas que los habitan afectan el desarrollo de las plantas de chile habanero y de sus frutos, y entre los suelos de Yucatán encontramos los suelos rojos y los suelos negros, los cuales tienen características fisicoquímicas y microbiológicas contrastantes (Medina-Lara *et al.*, 2019; Estrada-Medina *et al.*, 2016).

El proceso de desarrollo de las plantas y sus frutos está dictado por el genoma e influenciado por el ambiente que les rodea, afectando la expresión de los genes. La regulación de la expresión de los genes permite que ocurran los diferentes procesos biológicos y se sintetizen los compuestos que son parte de estas distintas etapas de desarrollo (Martínez-López *et al.*, 2014). Los capsaicinoides inician su síntesis desde etapas tempranas del desarrollo del fruto y se van acumulando, alcanzando su pico en etapas intermedias de madurez o en las etapas

INTRODUCCIÓN

más tempranas de la maduración completa, tras lo cual decae el contenido de estos metabolitos.

Las rutas metabólicas de los capsaicinoides incluyen la vía de los fenilpropanoides a partir de fenilalanina y la de síntesis de ácidos grasos de cadena ramificada a partir de leucina y valina, así como las rutas de síntesis de estos aminoácidos precursores (Arce-Rodríguez y Ochoa-Alejo, 2019). Se ha observado que la expresión de los genes relacionados con la ruta de síntesis de capsaicinoides está relacionada con el nivel de capsaicinoides encontrado en las distintas etapas de desarrollo de los frutos y regula la síntesis de estos compuestos en los frutos (Park *et al.*, 2019).

La expresión de los genes de estas vías se puede estudiar mediante la transcriptómica, que es el conjunto de transcritos que expresa una célula en cierto período y bajo ciertas condiciones (Arce-Rodríguez y Ochoa-Alejo, 2019). Las técnicas modernas de secuenciación han permitido el análisis masivo y en paralelo de genomas y transcriptomas, con rendimiento alto de múltiples muestras y pudiendo detectar aún genes que se expresan a niveles bajos. Estos análisis requieren herramientas bioinformáticas que permitan analizar la gran cantidad de datos transcriptómicos que finalmente a responder la pregunta de investigación (Kulski, 2016).

Este trabajo tiene como objetivo estudiar las diferencias de expresión de los genes de la ruta de síntesis de capsaicinoides en frutos de plantas de chile habanero crecidas en diferentes tipos de suelo de Yucatán, específicamente en suelos rojos y los suelos negros.

CAPÍTULO I

ANTECEDENTES

1.1. Género *Capsicum*. Generalidades

El género *Capsicum* incluye un grupo de especies de frutos picantes y no picantes pertenecientes a la familia de las solanáceas. Tanto a las plantas como a sus frutos se les denomina comúnmente como chiles. A este grupo pertenecen más de 30 especies, incluyendo especies domésticas como *Capsicum annuum*, *C. frutescens*, *C. chinense*, *C. baccatum* y *C. pubescens* (Tripodi y Kumar, 2019). Los chiles tienen un importante valor económico en varios países del mundo y se cultivan en diferentes países en todo el mundo.

Los frutos tienen propiedades sensoriales como sabor, olor y color que los hacen atractivos y muy usados en gastronomía. En los frutos se pueden encontrar carotenoides, flavonoides, compuestos volátiles, tocoferoles, capsaicinoides, entre otros nutrientes y compuestos con actividad benéfica para el organismo y que en conjunto confieren a los frutos sus llamativas características organolépticas. Son fuente de una gran variedad de metabolitos, incluyendo carotenoides, aminoácidos, vitaminas, compuestos con capacidad antioxidante y capsaicinoides (Antonio *et al.*, 2018; Aranha *et al.*, 2017).

1.2. El chile habanero (*Capsicum chinense* Jacq.)

El chile habanero es un cultivo tradicional de importancia económica en Yucatán, siendo una de las hortalizas más cultivadas en la Península (Ruiz-Lau *et al.*, 2011; Canto-Flick *et al.*, 2008). El chile habanero es una planta de ciclo anual. Su altura puede oscilar entre 75 y 120 cm en condiciones de invernadero. Su tallo es grueso, erecto y robusto. Las hojas son generalmente de color verde y de tamaño variable, simples, lisas, alternas y de forma lanceolada (Ruiz-Lau *et al.*, 2011).

Tiene una raíz principal tipo pivotante, que profundiza entre 0.40 a 1.20 m, con un sistema radicular bien desarrollado cuyo tamaño depende de factores como la edad de la planta y las características del suelo donde se desarrolle. La floración inicia cuando la planta empieza a ramificarse. Las flores son de color blanco y se presentan de manera individual o en grupos de dos o más en cada una de las axilas. Su tamaño varía entre 1.5-2.5 cm de diámetro de la

CAPÍTULO I

corola. El número de pétalos y sépalos es variable, de cinco a siete, al igual que la longitud del pedúnculo floral (Ruiz-Lau *et al.*, 2011). Las características del fruto pueden variar.

Generalmente es una baya hueca poco carnosa con forma acampanada que posee lóbulos. Su tamaño es de aproximadamente 6 cm de largo y 2.5 cm de ancho. En estado de inmadurez es de color verde, pero en estado de madurez presenta diferentes colores (morado, café, rojo, amarillo o naranja), de acuerdo a la variedad. Las paredes que dividen el interior del fruto son incompletas y en el extremo inferior se unen para formar unas estructuras membranosas que comúnmente se denominan venas, las cuales se insertan en la placenta, que es de color blanco amarillento y de apariencia esponjosa (González-Estrada *et al.*, 2018; Bosland y Votava, 2012). En la Tabla 1.1 se presenta la clasificación taxonómica del chile habanero.

Tabla 1.1. Clasificación taxonómica del chile habanero (*Capsicum chinense* Jacq.) (Bosland y Votava, 2012)

Reino	<i>Plantae</i>
División	<i>Magnoliophyta</i>
Clase	<i>Magnoliopsida</i>
Orden	<i>Solanales</i>
Familia	<i>Solanaceae</i>
Subfamilia	<i>Solanoideae</i>
Tribu	<i>Solaneae</i>
Subtribu	<i>Capsicinae</i>
Género	<i>Capsicum</i>
Especie	<i>Capsicum chinense</i>

1.3. Factores que influyen sobre las características fenotípicas del chile habanero

Los factores que influyen sobre las características fenotípicas del chile habanero son de tipo genético y ambiental, así como la interacción de ambos. Diferentes variedades o genotipos cultivados en las mismas condiciones presentarán fenotipos diferentes influenciados por su genoma y la expresión de los genes (Antonious *et al.*, 2009; Canto-Flick *et al.*, 2008). Por otro

lado, los factores ambientales causan que la misma variedad presente diferentes fenotipos (rendimiento del fruto, tiempo de madurez, número de frutos, contenido de metabolitos, entre otros). Esto es importante porque puede limitar un genotipo a determinados ambientes que le sean favorable para la producción de frutos, metabolitos, o en general, un fenotipo de interés (Jeeatid *et al.*, 2018a).

Debido a esto, en la Península de Yucatán, el chile habanero ha encontrado condiciones particulares para desarrollarse y originar distintas variedades con las características que le llevaron a ser distintos de aquellas variedades cultivadas en otras regiones del mundo, otorgándosele la Denominación de Origen de “Chile Habanero de la Península de Yucatán” (González-Estrada *et al.*, 2018). Por lo anterior, es importante la caracterización y conservación de los germoplasmas de chile habanero (Chhapekar *et al.*, 2020), pero también conocer la influencia de los factores ambientales sobre las características de las plantas y de sus frutos, así como la influencia que tienen por medio de la regulación genética (Peña-Yam *et al.*, 2019; Vaishnavi *et al.*, 2018).

Los factores medioambientales que afectan el desarrollo y producción del chile habanero incluyen el clima, la radiación solar, los tratamientos de fertilización, el régimen de riego y los factores fisicoquímicos y bióticos del suelo (Das *et al.*, 2016; Estrada-Medina *et al.*, 2016; Zamudio-Moreno *et al.*, 2014; Medina-Lara *et al.*, 2008). El chile habanero tiene un buen desarrollo en las regiones templadas y subtropicales.

Puede crecer en un rango de precipitación de 600-1250 mm y de temperatura de entre 10-35 °C, con una temperatura óptima de 30° C. La temperatura de germinación fluctúa entre 18-35 °C, siendo óptima a 30 °C. Los suelos más favorables para el desarrollo del chile habanero son aquellos que tienen un buen drenaje, una buena retención de humedad y un pH de 6.5-7.0, con una buena disponibilidad de nutrientes (Villa-Castorena *et al.*, 2014). El clima dominante de Yucatán es cálido subhúmedo con lluvias de verano. La temperatura promedio anual oscila entre 24.6 y 27.7 °C. El mes más frío es enero y el más cálido es mayo.

La precipitación pluvial oscila entre 415 y 1290 mm, distribuidos mayormente entre mayo y octubre. Al mismo tiempo, existe un gradiente de precipitación, encontrándose una tendencia al aumento de la precipitación anual del noroeste al oriente y sureste del estado. Por otro lado, la temperatura media anual tiende a aumentar en gradiente de noreste a noroeste-suroeste. De este modo, la parte norte y noroeste del estado tiende a tener menor humedad, menos lluvias y

más temporadas secas y calurosas en comparación con la parte oriente y sureste (Calzada-Rovirosa *et al.*, 2017; Orellana-Lanza *et al.*, 2010). En Yucatán, el chile habanero se suele cultivar por temporal, siendo destacada la temporada de lluvias (junio-septiembre), pues la temperatura, humedad y luminosidad favorecen los rendimientos. Sin embargo, está sujeto al ataque de plagas y enfermedades, especialmente a cielo abierto (González-Estrada *et al.*, 2018).

1.4. El género *Capsicum* y los capsaicinoides

El metabolismo puede definirse como el conjunto de transformaciones químicas organizadas en rutas que permiten el anabolismo y el catabolismo que tienen lugar en una célula u organismo. La mayor parte del carbono, nitrógeno y de la energía termina en las moléculas comunes a todas las células e incluyen carbohidratos, lípidos, aminoácidos, nucleótidos y sus derivados, y ciertos ácidos carboxílicos. Estos son los metabolitos primarios (Nelson y Cox, 2014).

Las plantas también destinan carbono y energía hacia la síntesis de moléculas orgánicas que no parecen tener relación directa con los procesos de fotosíntesis, respiración, asimilación de nutrientes, transporte de solutos o síntesis de macromoléculas. Éstos son los metabolitos secundarios y el conjunto de rutas involucradas es el metabolismo secundario y cumplen funciones ecológicas específicas (pigmentos o aromas, atracción de polinizadores, funciones protectoras contra depredadores o de defensa a patógenos) y a menudo su producción está restringida a un grupo específico de plantas (Ávalos-García y Pérez-Urria Carril, 2009).

En el género *Capsicum* se encuentran metabolitos secundarios como los carotenoides, flavonoides, compuestos volátiles y los capsaicinoides. Los compuestos volátiles incluyen terpenos, hidrocarburos, alcoholes, aldehídos, cetonas, aminas, ácidos, ésteres, lactonas y compuestos fenólicos. Algunos son responsables de los aromas en frutos y pueden tener actividad biológica. Los carotenoides son conocidos por aportan colores amarillos, naranjas y rojos a los frutos y tienen un papel importante en la fotosíntesis, la foto protección y estabilización de radicales libres. Los flavonoides, al igual que los carotenoides, tienen propiedades cromogénicas. También protegen contra la radiación UV, participan en la regulación del crecimiento, tienen actividad antimicrobiana y atraen polinizadores (Lemos *et al.*, 2019; Antonio *et al.*, 2018; Kim *et al.*, 2017).

Los capsaicinoides son un grupo de compuestos responsables de la sensación conocida como picor y son característicos de las plantas del género *Capsicum*. Tienen actividades biológicas y

aplicaciones industriales más allá de su interés gastronómico (Arce-Rodríguez y Ochoa-Alejo, 2019), incluyendo su uso como analgésicos, antiinflamatorios, antioxidantes, anticancerígenos, antitumorales, para beneficio del sistema gastrointestinal, termogénicos y en la regulación del peso corporal (Ghiasi *et al.*, 2019; Clark y Lee, 2016; Srinivasan, 2016; Caterina *et al.*, 1997). Otros usos de los capsaicinoides son como componentes de salsas y condimentos, de aerosoles pimienta para protección, de cosméticos, como repelentes de animales, además tienen propiedades antimicrobianas (Aza-González *et al.*, 2011).

Químicamente, los capsaicinoides se componen de un grupo de vanillilamina unido a una cadena alquílica por medio de un enlace amida (Figura 1.1). Se consideran pseudoalcaloides, con un nitrógeno extra cíclico en el grupo amida y son producto del metabolismo del aminoácido fenilalanina (de donde proviene el grupo vanillílico) y los aminoácidos de cadena ramificada como valina y leucina (de donde proviene el grupo acilo). Se consideran también como compuestos fenólicos derivados de la ruta de los fenilpropanoides. Los diferentes tipos de capsaicinoides se diferencian en la longitud, ramificaciones e insaturaciones de la cadena alquílica y pueden ser no saturados y ramificados, como en la capsaicina y la homocapsaicina, saturados y ramificados, como en la dihidrocapsaicina y la nordihidrocapsaicina, o saturados y lineales, como en la nonivamida (Antonio *et al.*, 2018) (Figura 1.1). Generalmente la capsaicina y la dihidrocapsaicina son los componentes mayoritarios de los capsaicinoides (aproximadamente el 90% del total), aunque las proporciones pueden variar según el estado de madurez y la variedad del chile (Benmansour-Gueddes, 2015).

El lugar de síntesis y acumulación de los capsaicinoides es en las células epidérmicas del septum interocular de la placenta, migrando del vesículas del retículo endoplásmico hacia las paredes celulares y formando vacuolas de almacenamiento en el espacio entre el exterior de las paredes celulares y la cutícula del septum (Tyler *et al.*, 2016; Zamski *et al.*, 1987), aunque también se han detectado en el pericarpio de algunas variedades muy picantes (Sugiyama, 2017). La acumulación de capsaicinoides ocurre desde etapas tempranas de desarrollo de los frutos y llegan a su máximo antes de la madurez o poco después de ésta, para después disminuir.

Esta síntesis y acumulación se da generalmente entre los 10-40 días post-antesis (DPA), aunque en algunas variedades puede continuar en etapas de desarrollo posteriores (Olguín-Rojas *et al.*, 2019). *C. chinense* produce frutos que contienen una mayor concentración de

capsaicinoides que otras especies de Chile (Sun *et al.*, 2022). El Chile habanero de la Península de Yucatán se encuentra entre los primeros lugares de picor (Canto-Flick *et al.*, 2008).

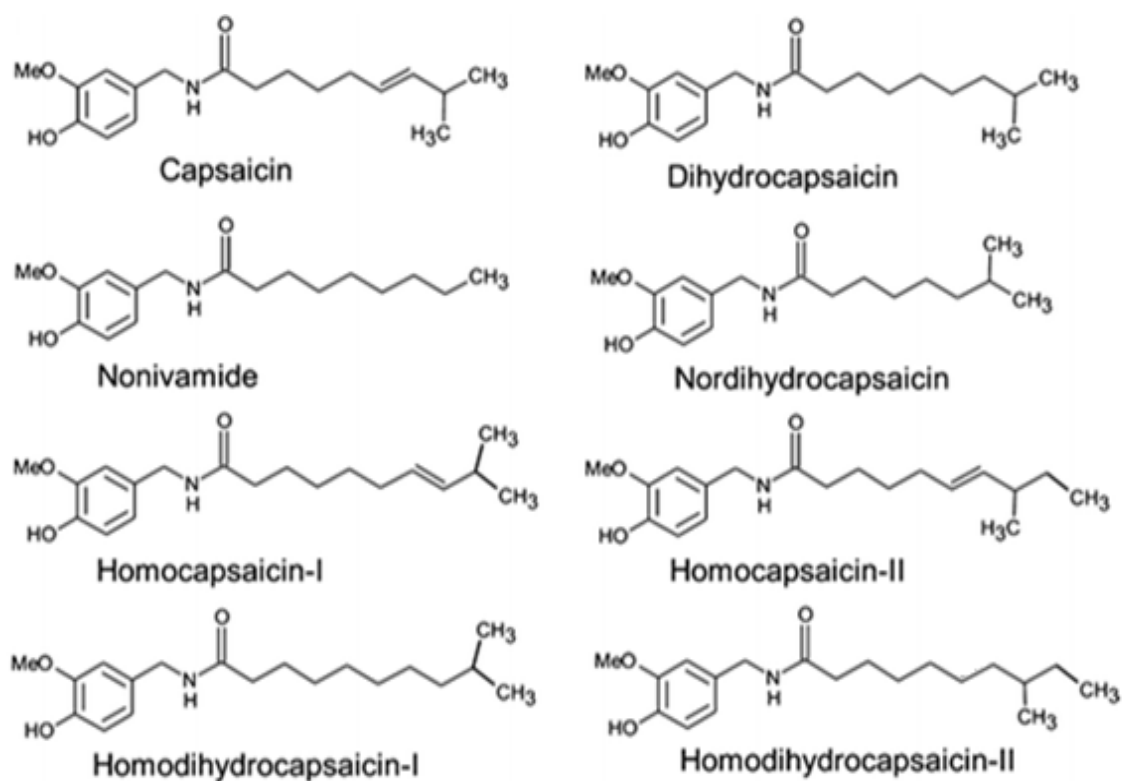


Figura 1.1. Estructura química de los capsaicinoides (Aza-González *et al.*, 2011)

La ruta metabólica de síntesis de capsaicinoides se compone de la conjunción de dos vías, la de los fenilpropanoides que da lugar a la vanillilamina a partir de la fenilalanina, y la síntesis de ácidos grasos de cadena ramificada que da lugar al grupo acilo, a partir de aminoácidos de cadena ramificada como la leucina y la valina (Figura 1.2) (Arce-Rodríguez y Ochoa-Alejo, 2019). La última reacción es una condensación de la vanillilamina con el grupo acilo derivado de los ácidos grasos catalizada por la capsaicinoide sintasa (CS) (Ogawa *et al.*, 2015).

Las rutas de síntesis de los aminoácidos precursores también son importantes para mantener la síntesis de capsaicinoides y se necesita una fuente de nitrógeno y de aminoácidos para proveer a los tejidos de éstos. Además, existe evidencia de que los frutos y sus tejidos placentarios tienen la capacidad enzimática para sintetizar dichos aminoácidos usando nitrógeno inorgánico y la inhibición de enzimas importantes de las rutas de síntesis de los

aminoácidos lleva a una disminución en el contenido de capsaicinoides en los frutos (Baas-Espinola *et al.*, 2016; Castro-Concha *et al.*, 2016).

Otra parte importante de la ruta de síntesis de capsaicinoides parece ser la síntesis del grupo acilo precursor, con una posible relación entre los capsaicinoides formados, los grupos acilo precursores y los aminoácidos precursores (Thiele *et al.*, 2008; Fujiwake *et al.*, 1980). Se ha encontrado una correspondencia entre el patrón del contenido de capsaicinoides y el patrón del contenido de los precursores tioésteres de ácidos grasos de cadena ramificada (acil-CoA y acil-ACP) en frutos de *C. annuum*, Jalapeño y *C. chinense*, Habanero, siendo los tioésteres más abundantes los correspondientes a los ácidos grasos precursores de la capsaicina (8-metil-trans-6-nonenamida) y a la dihidrocapsaicina (8-metil-nonanamida), los capsaicinoides más abundantes comúnmente encontrados en frutos (Thiele *et al.*, 2008).

También se ha observado que la valina es un precursor de capsaicinoides con un número par de átomos de carbono, como la capsaicina y la dihidrocapsaicina, y sus correspondientes ácidos grasos precursores; mientras que la leucina es un precursor de capsaicinoides con número impar de átomos de carbono, como la nordihidrocapsaicina y la homodihidrocapsaicina, y de los correspondientes ácidos grasos precursores (Suzuki *et al.*, 1981).

1.5. Factores que influyen en el contenido de capsaicinoides

La acumulación de los capsaicinoides en los frutos de chile depende del genotipo, el estado de desarrollo del fruto y los factores medioambientales, así como la interacción del genotipo y el medio ambiente (Arce-Rodríguez y Ochoa-Alejo, 2019; Jeeatid *et al.*, 2018a). Debido al genotipo, distintas variedades y especies tendrán diferente contenido de capsaicinoides. Como ejemplo de este factor genético, en 18 accesiones de *C. chinense* crecidas en invernadero bajo las mismas condiciones se determinó el nivel de picor, mostrando una diversidad intraespecífica del contenido de capsaicinoides.

En fruto entero la accesión NP1EG tuvo el mayor contenido de capsaicinoides. Sin embargo, en placenta el mayor contenido de capsaicinoides fue en la variedad NP4EC y en pericarpio fue la accesión NP3EC (Canto-Flick *et al.*, 2008). Igualmente, se ha reportado que, aunque en los frutos de chile habanero se sintetiza y se acumulan capsaicinoides en los tejidos de la placenta, en ciertas variedades se acumulan capsaicinoides también en el pericarpio, observándose altos niveles de expresión de genes de la ruta de síntesis de capsaicinoides,

como son *pAMT* (gen de la aminotransferasa de vanillilamina) y *PUN1* (gen de la CS), en el pericarpio y en la placenta de dichas variedades (Park *et al.*, 2019; Tanaka *et al.*, 2017).

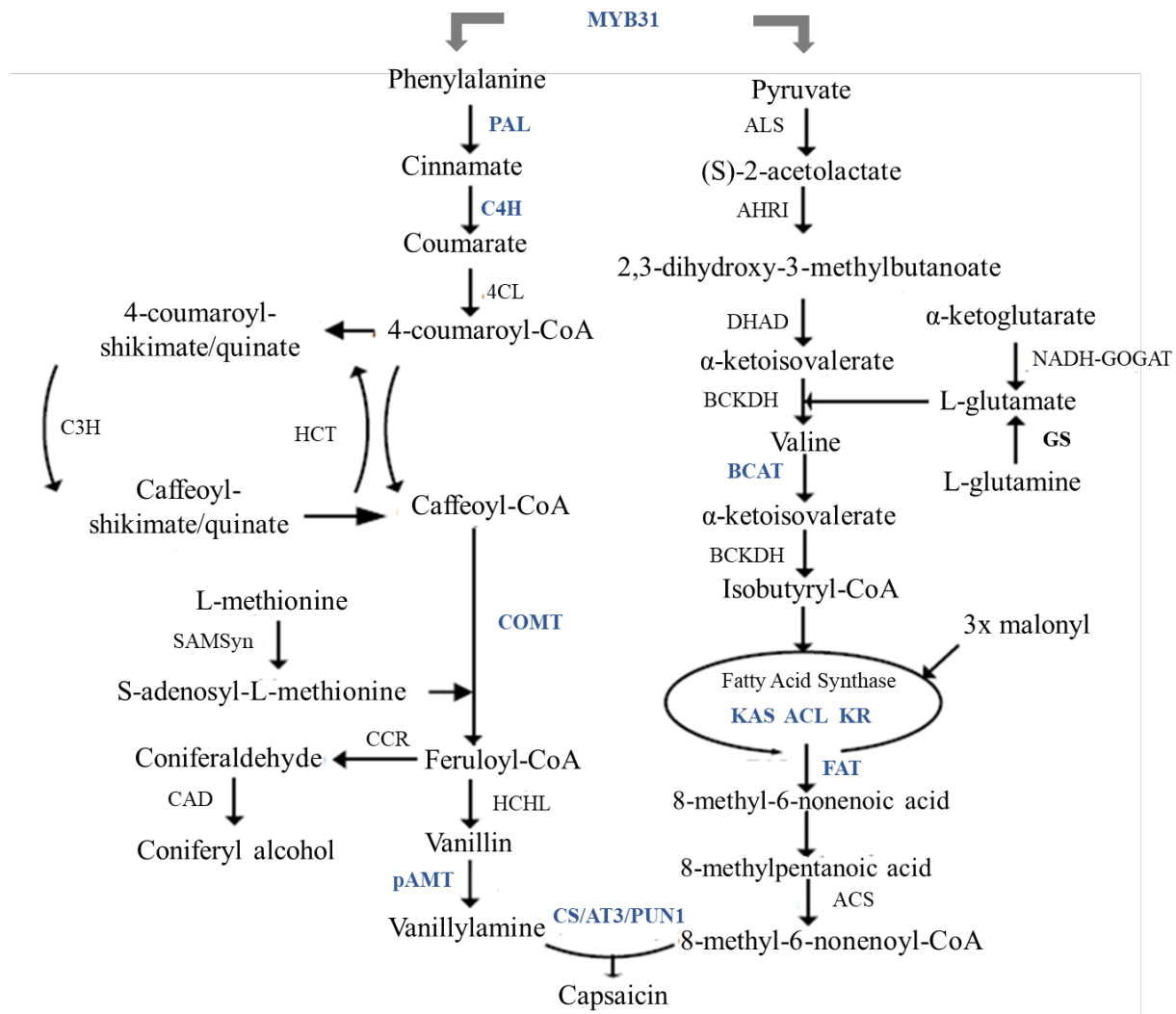


Figura 1.2. Ruta de biosíntesis de capsaicinoides. Las enzimas cuyos genes se han asociado con la presencia o no de picor o cuya expresión ha presentado consistencia o asociación con los cambios en la síntesis de capsaicinoides se señalan en azul. MYB31: factor de transcripción que tiene influencia en la expresión de genes de la ruta de los fenilpropanoides y de los ácidos grasos de cadena ramificada. PAL: fenilalanina amonio liasa; C4H: cinamato-4-hidroxilasa; 4CL: 4-cumaroil-CoA ligasa; CoA: coenzima A; HCT: hidroxicinamoil transferasa; C3H: cumaroil shikimate/quinate 3-hidroxilasa; COMT: ácido cafeico O-metiltransferasa; SAMSyn: S-adenosilmetionina sintetasa; HCHL: hidroxicinamoil-CoA hidratasa/liasa; CCR: cinamoil-CoA reductasa; CAD: cinamil alcohol deshidrogenasa; pAMT: aminotransferasa putativa; ALS: acetolactato α sintasa; AHRI: acetohidroxiácido reductoisomerasa; DHAD: dihidroxiácido deshidratasa; GS: glutamina sintetasa; NADH-GOGAT: glutamato sintasa dependiente de nicotinamida adenina dinucleótido; BCAT: transferasa de aminoácidos de cadena ramificada; BCKDH: deshidrogenasa/d Descarboxilasa de α -cetoácidos de cadena ramificada; KAS: cetoacil-ACP sintasa; ACP: transportadora de grupos acilo; ACL: proteína

transportadora de grupos acilo; CaKR1: cetoacil-ACP reductasa; FAT: acil-ACP tioesterasa; ACS: acil-CoA sintetasa; CS: capsaicinoide sintasa; AT3: aciltransferasa (Burgos-Valencia *et al.*, 2020)

Entre los factores ambientales que afectan el contenido de capsaicinoides, se pueden mencionar son la luz, la temperatura, el CO₂, la altitud, el clima, el suelo, el manejo de fertilización y el régimen de agua (Arce-Rodríguez y Ochoa-Alejo, 2019). El efecto de la interacción genotipo-ambiente fue estudiado por Jeeatid *et al.* (2018a) en frutos de nueve variedades de *C. chinense* crecidas en ambientes con distintas condiciones de temperatura del aire, humedad relativa e intensidad luminosa. La humedad relativa alta y la intensidad luminosa baja favorecieron un mayor contenido de capsaicinoides. Por otro lado, hubo variedades muy sensibles al cambio de medioambiente, tales como Phet Mordindang que tuvo el mayor contenido de capsaicinoides y rendimiento de capsaicinoides en condiciones muy específicas, mientras que otras variedades fueron menos sensibles al medio ambiente.

En el impacto de la fertilización, diferentes resultados se han observado de acuerdo al tratamiento con nitrógeno, la forma química de éste y a la variedad de chile a la cual se le aplica. Medina-Lara *et al.* (2008) reportaron en chile habanero que la fertilización con nitrógeno en forma de urea aumentó tanto el número de flores y frutos como el porcentaje de frutos/flores respecto a un control sin fertilización, siendo el tratamiento de 15 mM de urea el de mejor efecto. Los contenidos de capsaicina no aumentaron bajo fertilización con respecto al tratamiento control, y en algunos casos los contenidos de capsaicina disminuyeron, aunque el tratamiento con 15 mM de urea mantuvo el nivel de capsaicina con respecto al control.

Es decir, este tratamiento fue eficiente para mantener los niveles de capsaicina en los frutos de las plantas, y al mismo tiempo aumentó la producción de flores y frutos con respecto a un tratamiento sin fertilización. De manera similar, Monforte-González *et al.* (2010) observaron un aumento en el contenido de capsaicinoides en frutos y placentas cuando se aumentó la dosis de nitrato aplicada en plantas y el contenido de nitrógeno en placenta tuvo concordancia con el contenido de capsaicinoides. En otro trabajo, la aplicación de nitrógeno en diferentes relaciones diferentes de amonio:nitrato (0:100, 25:75 y 50:50) tuvo un efecto sobre el contenido de capsaicina y dihidrocapsaicina en placenta, las actividades de enzimas de la ruta de biosíntesis de capsaicinoides (PAL, CS, GOGAT y GS) y la expresión de genes relacionados (*AT3*, *NADH-GOGAT*, *GS*, *FAT*, *COMT*) (Zhang *et al.*, 2020a).

El régimen de riego es otro factor importante en el contenido de capsaicinoides, como se puede observar en el trabajo de Zamudio-Moreno *et al.* (2014), donde se demostró que hay una modificación en los contenidos de capsaicinoides ante varios tratamientos de riego sobre plantas de chile habanero (riego diario o control, riego cada siete días o R7 y riego cada nueve días o R9) que también tuvo un efecto sobre las actividades enzimáticas de CS (síntesis de capsaicinoides) y las peroxidasa de capsaicinoides (catabolismo de capsaicinoides), influyendo por lo tanto en la formación, mantenimiento y disminución en el contenido de capsaicinoides en los frutos. Aunque se ha observado que el riego cada cierto tiempo favorece la síntesis de capsaicinoides con respecto a un tratamiento de riego diario, el efecto del régimen de riego es distinto entre variedades de plantas (Jeeatid *et al.*, 2018b).

La temperatura también es un factor que tiene efecto sobre el contenido de capsaicinoides en frutos, y su efecto difiere entre variedades, pudiendo ocurrir un aumento en el contenido de estos metabolitos a altas temperaturas (40-48 °C), o bien, estar favorecidos a temperaturas menores (33-40 °C) (González-Zamora *et al.*, 2013).

El tipo de luz aplicada también ha mostrado tener un efecto sobre el contenido de capsaicinoides en frutos de plantas de chile. Como ejemplo de este factor, en *C. annuum* L cv. Cheonyang la luz LED azul aumentó el contenido de capsaicinoides en los frutos en comparación con las plantas expuestas a luz LED roja, roja con azul o luz fluorescente (Gangadhar *et al.*, 2012).

1.6. Los suelos cultivables en Yucatán

En Yucatán predominan los suelos tipo leptosol. Estos suelos incluyen suelos muy delgados sobre roca continua y suelos que son extremadamente ricos en fragmentos gruesos. Su material parental es de varios tipos de roca continua o de materiales no consolidados con menos del 20% en volumen de tierra fina. Estos suelos también pueden presentar dificultades en la agricultura por su porosidad, baja profundidad y pedregosidad, limitando el crecimiento de las raíces y causando sequía incluso en ambientes húmedos (FAO, 2015).

El estado de Yucatán tiene un extenso territorio de topografía ondulada, formando una planicie kárstica desarrollada sobre capas de piedra caliza cenozoica. El lecho rocoso calcáreo y el clima subhúmedo han contribuido a la formación de suelos con diferentes propiedades fisicoquímicas como profundidad y hierro, aluminio como material residual. Además, los suelos tienen una distribuidos en mosaico. En el norte del estado la mayoría de los suelos tienen una

profundidad de hasta 25 cm. Algunos presentan un lecho de roca calcárea continua a los 10 cm de profundidad (leptosoles líticos), mientras que otros leptosoles tienen una profundidad de entre 10-25 cm (Bautista *et al.*, 2005).

Una forma muy común de clasificar los suelos de Yucatán es mediante el sistema maya, basado en características observables y en la experiencia en la agricultura de los pueblos tradicionales que habitan la región. Bajo esta clasificación se han realizado investigaciones para identificar las propiedades fisicoquímicas de estos suelos. También se han encontrado tendencias en la distribución de estos suelos (Estrada-Medina *et al.*, 2013; Bautista *et al.*, 2005; Bautista-Zúñiga *et al.*, 2003).

Entre los tipos de suelo en Yucatán se encuentran los suelos rojos (*Haylu'um*), suelos café-rojizos (*Chaclu'um*) y suelos negros (*Boxlu'um*), cada uno con sus características fisicoquímicas. Además, estos tipos de suelos suelen diferenciarse en su localización en el relieve yucateco. Los suelos negros tienden a ubicarse en los montículos o elevaciones, los suelos cafés en el pie de monte y suelos rojos en las partes bajas, planicies o depresiones (Bautista-Zúñiga *et al.*, 2003).

Ciertas propiedades fisicoquímicas como la textura, la densidad aparente, la materia orgánica, el nitrato, el fósforo y el calcio también pueden ser relacionadas con el color del suelo. En estudios de suelos en el norte de Yucatán se encontró que el contenido de materia orgánica, nitrato y calcio fue mayor en los suelos elevados (suelos negros) y disminuyó conforme se descendía la altura del suelo (suelos rojos), mientras que los suelos café-rojizos ocuparon una posición intermedia. Los suelos rojos o café-rojizos también tendieron a tener mayor densidad aparente, mayor contenido de partículas más finas (arcilla y limo) que los suelos negros y estos últimos tendieron a tener mayor contenido de arena (Bautista *et al.*, 2005; Bautista-Zúñiga *et al.*, 2003). El contenido de materia orgánica, arena, la capacidad de intercambio de cationes, contenido de carbonatos, pH y potasio fue mayor en suelos negros de los montículos que en suelos rojos de las planicies, que tuvieron mayor contenido de arcilla y limo, en Hocabá, Yucatán (Estrada-Medina *et al.*, 2013).

Estrada-Medina *et al.* (2016) realizaron una caracterización microbiológica de suelos negros y rojos de Yucatán. En características fisicoquímicas, el suelo negro tuvo un mayor contenido de arena, nitrógeno, fósforo, carbono orgánico, carbonatos, materia orgánica, capacidad de campo, conductividad eléctrica y potencial redox; mientras que el suelo rojo tuvo un mayor

porcentaje de arcilla. Las poblaciones de bacterias y hongos también fueron diferentes. En el suelo negro la población bacteriana más abundante fue el phylum de Actinobacteria (43.7%), seguido de Acidobacteria (26.9%) y Proteobacteria (23.6%); y en suelo rojo dominó el grupo Acidobacteria (83%). El grupo fúngico más abundantes en suelo negro fue Ascomycota (77%); mientras que en suelo rojo fueron los grupos Ascomycota (56%) y Basidiomycota (22%).

Las diferencias entre suelos influyen de manera importante sobre la forma química de los nutrientes, su movilidad y captación por las plantas. Por ejemplo, los suelos de diferentes regiones de Yucatán tienen distinto contenido de potasio soluble y potasio intercambiable, difusión efectiva y poder amortiguador (capacidad de mantener una concentración de potasio en la solución del suelo) (Borges-Gómez *et al.*, 2005). Medina-Lara *et al.* (2019) estudiaron la influencia de suelos rojo, café y negro irrigados con agua de alta conductividad sobre el desarrollo y el contenido de nutrientes de plantas de chile habanero.

En estas condiciones, los suelos rojo y café fueron los mejores para la producción y desarrollo de los frutos. El suelo negro tuvo alta conductividad eléctrica y gran retención y acumulación de sodio del agua usada para la irrigación, además de que tuvo alta retención de nitrógeno, fósforo y potasio, mientras que las plantas crecidas en suelos rojo y café pudieron tomar y llevar los nutrientes a sus tejidos (raíz y hojas) más eficientemente

El tipo suelo también tiene un efecto sobre contenido de capsaicinoides. Esto es, suelos con características fisicoquímicas y biológicas diferentes, (contenido de materia orgánica, nitrógeno, fósforo, textura, densidad aparente, capacidad de retención de agua y actividad microbiana) tienen una producción diferente de capsaicinoides (Das *et al.*, 2016).

En suelos negro y rojo de Yucatán se ha encontrado un mayor contenido de capsaicinoides en los frutos en distinto estado de madurez (inmaduros, madurez incompleta y maduros) de plantas cultivadas en suelo rojo, mientras que los frutos inmaduros y de madurez incompleta en suelo negro y los frutos maduros en suelo café tuvieron el menor contenido de capsaicinoides al comparar tipos de suelo en frutos en la misma etapa de desarrollo (Rodríguez-Buenfil *et al.*, 2018).

1.7. Expresión de genes y transcriptómica de la ruta de biosíntesis de capsaicinoides

La transcripción es un proceso que produce una secuencia de RNA correspondiente a una secuencia de DNA (gen). Es el primer paso en el proceso de expresión de los genes y

frecuentemente es muy regulado. En algunos casos el transcrito es usado para la producción de un polipéptido, por lo que se trata de RNA mensajero (mRNA), y en otros el transcrito es el producto génico funcional, como los RNA ribosomales (rRNA) y los RNA de transferencia (tRNA) (Krebs *et al.*, 2018). El transcriptoma es el conjunto de transcritos de un organismo bajo condiciones específicas y en cierto momento espaciotemporal (Arce-Rodríguez y Ochoa-Alejo, 2019).

Las tecnologías de secuenciación de nueva generación (next-generation sequencing, NSG) hacen posible secuenciar el transcriptoma completo mediante RNA-Seq. Esta tecnología puede ser descrita como un conjunto de metodologías o protocolos para obtener información de un conjunto de moléculas de RNA de una muestra biológica, determinando la secuencia primaria y la abundancia relativa de cada RNA (Liu *et al.*, 2013). Las moléculas de cDNA o RNA se cortan en fragmentos pequeños, que se secuencian posteriormente. Se crea una biblioteca ligando adaptadores, tags, primers en los extremos de los fragmentos. Se hace la amplificación de la biblioteca en superficies sólidas o cuentas donde existen emulsiones en donde ocurre la secuenciación (Kulski, 2016).

La incorporación del nucleótido se realiza directamente por luminiscencia o por cambios en una carga eléctrica durante el procedimiento de secuenciación. Los pequeños fragmentos de secuencias, denominados “*reads*” (lecturas), deben ser unidos para reconstruir el transcriptoma original, proceso que se denomina ensamblaje. Las estrategias de ensamblaje del transcriptoma pueden ser ensamblaje por comparación, usando un transcriptoma o genoma de referencia, y ensamblaje *de novo*, donde se aprovecha la redundancia entre fragmentos contiguos para unirlos y formar fragmentos más grandes hasta tener la secuencia completa (Kulski, 2016; Martin y Wang, 2011).

En *Capsicum* se han realizado estudios acerca de expresión y transcriptómica de los genes de la ruta de biosíntesis de capsaicinoides. Uno de los descubrimientos que se ha realizado es que ciertas variedades de plantas del género *Capsicum* no picantes tienen alelos no funcionales de genes de la ruta de biosíntesis, entre los que están genes de la CS, la pAMT, KR y factores de transcripción, principalmente de tipo MYB (Egan *et al.*, 2019; Han *et al.*, 2019; Koeda *et al.*, 2019; Tsurumaki y Sasanuma, 2019).

El gen de la CS, la enzima del último paso de la ruta de síntesis de capsaicinoides (conocido como *AT3* o *PUN1*), es uno de los genes que controla de manera importante el picor en

Capsicum. Muchas variedades no picantes de *C. chinense*, *C. annuum* y *C. frutescens* presentan alelos no funcionales (Kirii *et al.*, 2017; Stellari *et al.*, 2010; Stewart Jr. *et al.*, 2007; Stewart Jr. *et al.*, 2005). Además, se ha encontrado en el genoma de *Capsicum* varias copias del gen *AT3* funcionales y no funcionales, ubicadas en el cromosoma 2 y que se diferencian por varias sustituciones e inserciones/deleciones e incluso recombinaciones, además de que posiblemente no todas las copias tengan la misma importancia en el control de la síntesis de los capsaicinoides (Egan *et al.*, 2019; Qin *et al.*, 2014).

Sobre el gen *pAMT* (que codifica la aminotransferasa que cataliza la transaminación de la vanillina para formar vanillilamina), al igual que el gen *AT3*, tiene un control sobre el picor en frutos de plantas de chile y se han encontrado varios alelos no funcionales en varias variedades no picantes de *C. annuum*, *C. frutescens* y *C. chinense* (Tsurumaki y Sasanuma, 2019; Tanaka *et al.*, 2018; Park *et al.*, 2015a).

Varios factores de transcripción también han mostrado tener un control sobre la expresión de los genes de la ruta de síntesis de capsaicinoides y sobre el contenido de estos metabolitos en frutos. Durante el desarrollo de frutos de *C. annuum* se demostró la coexpresión de genes de factores de expresión R2R3-MYB y genes de la ruta biosintética de capsaicinoides (Wang *et al.*, 2020).

Estudios de expresión reportan en *C. annuum* 'Tampiqueño74' un gen de un factor de transcripción MYB, *CaMYB31*, cuyos niveles de expresión en tejidos de placentas tuvieron un comportamiento similar con los cambios en la acumulación de capsaicina y dihidrocapsaicina durante el desarrollo del fruto. Al contrario, la variedad no pungente 'California Wonder' tiene una expresión muy baja o nula en las mismas etapas de desarrollo (Arce-Rodríguez y Ochoa-Alejo, 2017). El gen del factor de transcripción *CaMYB31* fue localizado en el cromosoma 7 de *C. annuum*, donde una variedad no pungente (YCM334) presentó una mutación sin sentido en su primer exón que llevó a un codón de parada prematuro. Dicha mutación no estuvo presente en la variedad pungente 'Tean' y se relacionó con el control del picor. Además, genes de la ruta de síntesis de capsaicinoides, como *PUN1*, *KAS*, *BCAT* y *FAT*, tuvieron una expresión baja en la variedad no picante en comparación a la variedad picante en frutos entre los 26 y 36 DPA (Han *et al.*, 2019). Se ha mostrado que este activador transcripcional actúa uniéndose al promotor de genes de la ruta de síntesis de capsaicinoides, como *AT3* y regula positivamente la expresión de éstos (Zhu *et al.*, 2019). Además, este factor de transcripción es fuertemente expresado no sólo en placenta, sino también en pericarpio de variedades muy picantes de *C.*

chinense, como 'Trinidad Moruga Scorpion', y tiene un control sobre la expresión de genes de la ruta de síntesis de capsaicinoides en los pericarpios de estas variedades (Sun *et al.*, 2022).

Estudios transcriptómicos en *C. annuum* 'Tampiqueño', mediante RNA-Seq, han mostrado que el patrón de expresión de genes de la ruta de síntesis de capsaicinoides está relacionado con los cambios en el contenido de capsaicinoides en frutos durante su desarrollo. Esto es un incremento de 10 a 20 DPA en la expresión, seguido de una disminución de la entre los 20 y 40 DPA y de 40 a 60 DPA. Entre estos genes estuvieron *PUN1*, *COMT*, *PAL*, *FAT*, *ACL*, *pAMT*, *KAS*, *BCAT* y *C4H* (Martínez-López *et al.*, 2014). Otros estudios comparativos de la expresión de genes entre variedades picantes y no picantes y entre estados de desarrollo de frutos muestran que genes como *PUN1*, *pAMT*, *KAS*, *BCAT*, *PAL*, *C4H*, *ACS*, *NADH-GOGAT*, *COMT* y *BCKDH* tienen patrones de expresión que concuerdan con los cambios en el contenido de capsaicinoides y que su nivel de expresión es mayor en las variedades picantes que en las no picantes (Zhang *et al.*, 2016; Kim *et al.*, 2014).

Los factores ambientales también afectan la expresión de genes de la ruta de síntesis de capsaicinoides. El promotor de *PUN1* tiene motivos y secuencias consenso de respuesta a luz, estrés térmico y mecánico (Kim *et al.*, 2009). Puede inducirse por factores como daño mecánico, calor y precursores de la ruta de biosíntesis de capsaicinoides (fenilalanina y valina). Además, la expresión es diferencial en distintos tejidos de la planta, teniéndose evidencia de que la expresión de los genes es más fuerte en frutos y tejidos reproductivos en comparación de raíz y hojas (Kirke *et al.*, 2018; Kim *et al.*, 2009).

La expresión de genes como los de factores de transcripción *MYB*, *KAS* y *pAMT*, pueden verse afectadas por hormonas como el ácido salicílico, ácido indolacético, jasmonato de metilo y por el daño mecánico, la luz, la sequía, estrés salino y la temperatura, y el efecto es diferente entre variedades y entre genes (Sun *et al.*, 2019; Arce-Rodríguez y Ochoa-Alejo, 2017). En cultivos *in vitro* de placentas el ácido jasmónico, el metiljasmonato, el ácido salicílico y el quitosano han estimulado el aumento de expresión de genes de la síntesis de capsaicinoides (*PUN1*, *pAMT*, *COMT*, *C3H*, *HCT*, *PAL*, *NADH-GOGAT*) y la acumulación de capsaicina y dihidrocapsaicina (Kabita *et al.*, 2020).

También se ha observado que la fuente de nitrógeno tiene un efecto sobre la expresión de genes de la ruta de biosíntesis de capsaicinoides. En aplicaciones de amonio:nitrato, la relación entre éstos tiene un efecto sobre el contenido de capsaicinoides, las actividades de las enzimas

CAPÍTULO I

CS, PAL, GOGAT y GS, y los genes *NADH-GOGAT*, *GS*, *AT3*, *FAT*, *BCAT*, *C4H*, *COMT* y *PAL* en frutos de *C. annuum* (Zhang *et al.*, 2020a).

Con los antecedentes anteriores queda clara la importancia del conocimiento de la expresión de los genes relacionados con la síntesis de capsaicinoides, especialmente de aquellos que controlan el picor en los frutos. Pese a esto, se conoce poco sobre cómo afectan los estímulos ambientales la expresión de los genes relacionados con la biosíntesis de capsaicinoides, y, particularmente en suelos, donde coexisten varios factores fisicoquímicos y biológicos. En el chile habanero a la fecha existe poca información sobre transcriptoma en plantas crecidas en distintas condiciones en el estado de Yucatán. Además, se sabe que los suelos negros y rojos tienen características fisicoquímicas contrastantes y que los frutos de plantas de chile habanero crecidas en esas distintas condiciones presentan diferencias en su fenotipo, incluyendo el contenido de capsaicinoides que está relacionado con el picor.

Sabiendo que esta característica es muy apreciada en el chile habanero como alimento, así como los variados usos que tienen los capsaicinoides en la industria, en este trabajo se propone realizar un estudio comparativo del transcriptoma de frutos de chile habanero crecido en suelo negro y suelo rojo para conocer las diferencias de expresión de genes de las rutas metabólicas relacionadas con la biosíntesis de capsaicinoides, y como se podrían relacionar las diferencias de expresión con las diferencias observadas en el contenido de capsaicinoides de los frutos.

PREGUNTA DE INVESTIGACIÓN

¿Tiene el tipo de suelo en el que se desarrollan las plantas de chile habanero un efecto diferencial sobre los niveles de expresión de los genes relacionados con el metabolismo de capsaicinoides en los frutos?

HIPÓTESIS

Si el tipo de suelo en el que crece la planta está involucrado en la modificación del contenido de capsaicinoides en frutos de chile habanero, éste tendrá un efecto diferencial sobre la expresión de genes relacionados con el metabolismo de capsaicinoides en los frutos.

JUSTIFICACIÓN

El chile habanero es una de las plantas con más importancia cultural y económica de Yucatán, siendo la segunda hortaliza más cultivada en el estado, después del tomate. Los frutos del chile habanero producen una gran variedad de metabolitos secundarios entre los cuales se encuentran los capsaicinoides que imparten a los frutos la sensación de picor, atributo apreciable por sus consumidores. Los capsaicinoides además tienen efectos y usos farmacológicos y antimicrobianos, repelentes de animales o en la fabricación de aerosoles. Siendo estos metabolitos de interés por su variedad de aplicaciones, es deseable un mejor entendimiento de las condiciones medioambientales que controlan su producción en las plantas de chile habanero para tener bases que permitan diseñar estrategias que lleven a mejorar las características deseables en una planta, como el contenido de los capsaicinoides. El suelo es un factor importante en el control del metabolismo de las plantas.

En Yucatán, existen varios suelos con características diferentes y que influyen en la calidad de una planta y sus frutos (Medina-Lara *et al.*, 2019). Para un mejor entendimiento de cómo influye el suelo en la síntesis de capsaicinoides en los frutos y que tipo de suelo es más favorable para la síntesis de capsaicinoides, se puede estudiar que genes gobiernan o influyen en las diferencias en el contenido de capsaicinoides en frutos de plantas crecidas en suelos con características contrastantes (suelo negro y suelo rojo). El conocimiento de la dinámica de la expresión genética y de la regulación de estos genes es de interés ya que permite tener una base científica para controlar la producción de capsaicinoides en los frutos lo cual es parte de la calidad de los mismos.

OBJETIVO GENERAL

Estudiar la expresión de genes relacionados al metabolismo de capsaicinoides en frutos de plantas de chile habanero (*Capsicum chinense* Jacq.) cultivadas en diferentes tipos de suelos (suelo negro y suelo rojo)

OBJETIVOS ESPECÍFICOS

- Evaluar el efecto del tipo de suelo sobre los parámetros de floración, fructificación y el contenido de capsaicinoides en plantas cultivadas en suelo negro y suelo rojo
- Comparar la expresión de genes relacionados con el metabolismo de capsaicinoides en frutos de plantas cultivadas en los diferentes suelos
- Analizar bioinformáticamente las regiones reguladoras, filogenia, y la afinidad por docking hacia sustratos de un gen relacionado con el metabolismo de capsaicinoides expresado en los frutos de plantas cultivadas en los distintos suelos

ESTRATEGIA EXPERIMENTAL

Las plantas utilizadas pertenecen a chile habanero variedad Jaguar. Se trasplantaron 15 plántulas en suelo negro y 15 en suelo rojo, en macetas con 12 kg de suelo. Se realizó un ciclo de cultivo de 150 días posttrasplante (DPT). Durante este ciclo se etiquetaron diariamente las flores en día de antesis y se recolectaron diariamente frutos de 45 días DPA, en los cuales se midió el peso, largo y ancho de los frutos. También se recolectaron frutos de 25, 45 y 60 DPA. Se seleccionaron frutos de 25 DPA para hacer dos grupos de 5 frutos por tipo de suelo. Se extrajo el RNA total de cada grupo y se realizó la secuenciación para obtener el transcriptoma, mandando las muestras a Macrogen, Corea del Sur. Se realizó un ensamblado de *novο* para obtener la biblioteca de *contigs*, de donde se realizó un análisis de enriquecimiento de rutas metabólicas con las anotaciones KEGG (Kyoto Encyclopedia of Genes and Genomes) comparando las expresiones de los genes entre tipos de suelo, y se analizó la expresión de genes de las rutas enriquecidas y de genes de la ruta de síntesis de capsaicinoides en los frutos de 25 DPA entre suelo negro y suelo rojo.

De entre los genes, se eligió el gen de la *AT3* para realizar un análisis de las características del gen y de los distintos haplotipos, así como un análisis filogenético de secuencias, buscando secuencias homólogas en la base de datos del NCBI. Además, de los distintos grupos de haplotipos se realizó un análisis de las secuencias de reguladores, y el acoplamiento o 'docking' con diferentes sustratos de la enzima CS, modelando previamente la estructura tridimensional de ésta para cada grupo de haplotipos. Finalmente, se realizó un análisis genómico de los probables factores de transcripción en *CIS* de los haplotipos CS identificados con el fin de relacionar las diferencias encontradas entre los haplotipos y su expresión en suelos con características contrastantes.

Un esquema del diseño experimental se muestra en la Figura 1.3.

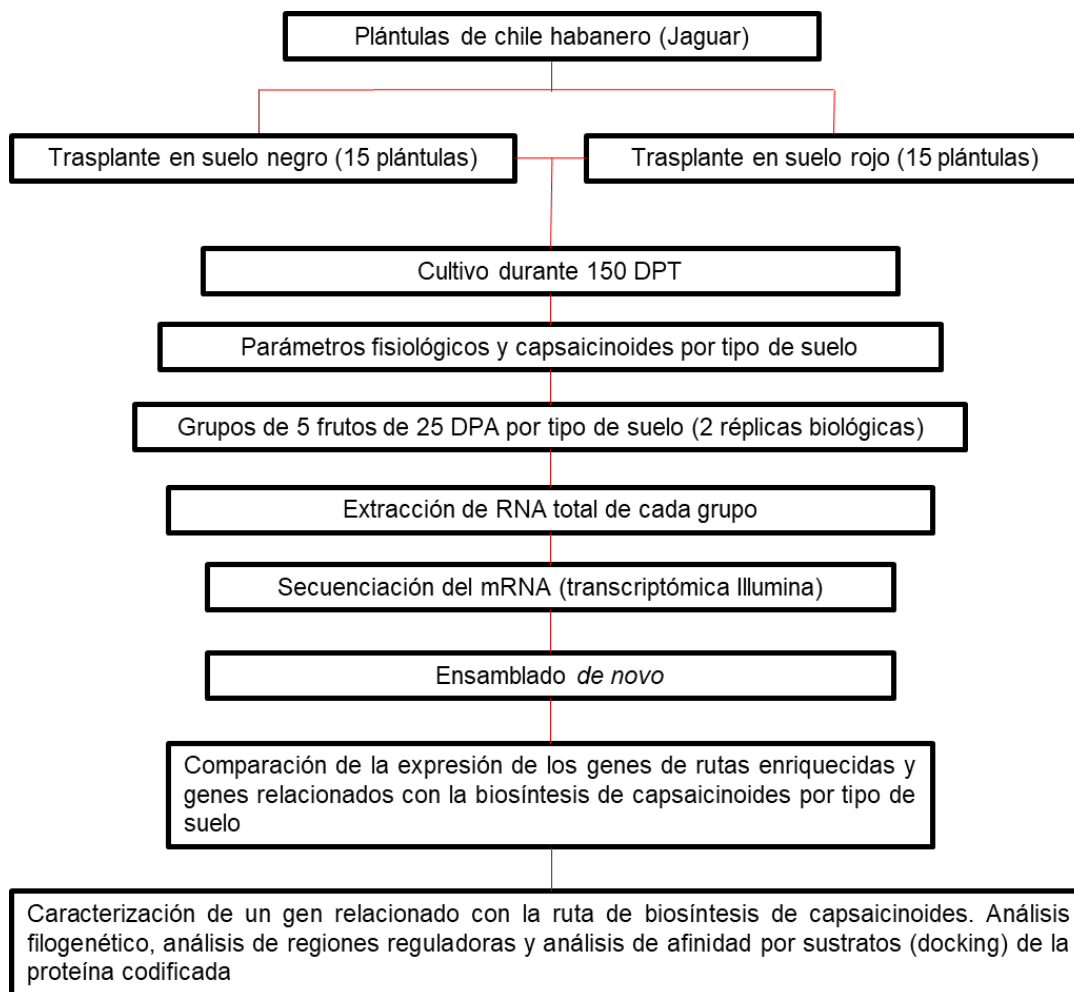


Figura 1.3. Esquema de la estrategia experimental

CAPÍTULO II

Artículo publicado en Brazilian Journal of Botany (2020) 43:201–212,

<https://doi.org/10.1007/s40415-019-00575-6>

GENE EXPRESSION RELATED TO THE CAPSAICINOIDS BIOSYNTHESIS IN THE *Capsicum* GENUS: A MOLECULAR AND TRANSCRIPTOMIC APPROACHES

Burgos- Valencia, Eduardo¹, Echevarría-Machado, Ileana¹, Narváez-Zapata, José A.², Martínez-Estévez, Manuel^{1*}

ABSTRACT

A review of the gene expression related with capsaicinoid biosynthesis in different species and varieties within the *Capsicum* genus is carried out. The results obtained by different molecular approaches is described in different species of *Capsicum* (mainly *C. frutescens*, *C. chinense* and *C. annuum*). The analysis of these results indicate that some genes determine the pungency or non-pungency in varieties or species, mainly by recessive or non-functional alleles. In addition, the level of expression of other genes modulate the level of capsaicinoids. Environmental factors and abiotic stress influence the expression of genes candidates to change the capsaicinoid biosynthesis and the fruit pungency. Other factors like fruit tissue (placenta, pericarp) and plant variety also influence the transcriptomic profile. Finally, transcriptome approaches show a gene regulation in the capsaicinoid biosynthesis pathway, and interesting, in other genes covering the phenylpropanoids synthesis and the amino acid precursor synthesis that have roles in the capsaicinoids accumulation.

Keywords: habanero pepper, gene expression, transcriptomic profile

2.1. INTRODUCTION

The genus *Capsicum* is a group of plants of the Solanaceae family and related with other crops as tomato, petunia and potato. Several species have been domesticated within the *Capsicum* genus, including *C. annuum*, *C. baccatum*, *C. frutescens*, *C. pubescens* and *C. chinense* Jacq. (Tripodi and Kumar, 2019; Bosland and Votava, 2012; Ruiz-Lau *et al.*, 2011). Its fruits have a great variety of metabolites of industrial interest (Baenas *et al.*, 2019; Antonio *et al.*, 2018; Kim

et al., 2017) that give them its appreciated sensory (Guzmán and Bosland, 2017; Bogusz Junior *et al.*, 2015; Eggink *et al.*, 2012) and its bioactive properties (Di Sotto *et al.*, 2018; Sosa-Moguel *et al.*, 2018; Mokhtar *et al.*, 2017). Among all metabolites, capsaicinoids are of great interest due they give the fruits its characteristic spiciness. Capsaicinoids have a variety of bioactive properties like analgesic, anti-inflammatory, antioxidant, anticancer and antitumor (Ghiasi *et al.*, 2019; Lin *et al.*, 2018; Clark and Lee, 2016; Srinivasan, 2016; Caterina *et al.*, 1997) These specific characteristics are due in part to the environmental conditions in which the plants grow. The genotype, the environment and the interaction of these factors influence the development of the fruit and its chemical content, including the capsaicinoids accumulation. Capsaicinoid accumulation also depends on the development stage of the fruit (Arce-Rodríguez and Ochoa-Alejo, 2019; Jeeatid *et al.*, 2018a; Ruiz-Lau *et al.*, 2011). Genotype influence implies that different *Capsicum* species or different accessions or varieties of the same species will have different phenotype and chemical content (Canto-Flick *et al.*, 2008). The environment also must be considered since a same variety will present different phenotypes when it grows and it develops in different locations or even at different times of the year. Environment factors that influence the content of capsaicinoids includes the light, temperature, carbon dioxide, altitude, soil, precipitation, relative humidity, fertilization and water managements (Arce-Rodríguez and Ochoa-Alejo, 2019; Jeeatid *et al.*, 2018a; Das *et al.*, 2016; Latournerie-Moreno *et al.*, 2015; Zamudio-Moreno *et al.*, 2014; González-Zamora *et al.*, 2013; Gangadhar *et al.*, 2012; Medina-Lara *et al.*, 2008). The capsaicinoids begin their synthesis in early and intermediate stages of fruit development peaking in the last stages before the fruits ripen, and in the earliest stages of ripening, after which the content of these metabolites decreases (Castro-Concha *et al.*, 2016). The metabolic pathways of capsaicinoids include the phenylpropanoid pathway, the branched-chain fatty acid pathway synthesis, the branched-chain amino acid pathway synthesis and the shikimate pathway and phenylalanine synthesis (Naves *et al.*, 2019; Zhang *et al.*, 2016). The model of this pathway is shown in Figure 2.1.

The differences in the content of capsaicinoids among *Capsicum* plants can be attributed at differences on expression of the genes of the capsaicinoid pathway, with depends on the genotype and the influence of the environment. The expression of the genes of this pathway can be investigated by transcriptomics, which is responsible for studying the transcriptome, that is, the set of transcripts that a cell expresses in a certain period and under certain conditions. Understanding the transcriptome is essential to interpret the functional elements of the genome and how occurs the expression of the genes (Liu *et al.*, 2013; Liu *et al.*, 2012). The

transcriptome can be studied experimentally *in vivo* and computationally *in silico*. Modern sequencing techniques have allowed the massive and parallel analysis of genomes and transcriptomes, with high performance of multiple samples and can still detect genes that are expressed at low levels. However, these sequences occur in small fragments, so in order to perform these analyzes is necessary to fragment the transcriptome and reassemble the sequences once obtained. For the analysis of genomes and transcriptomes, bioinformatic tools are required to reassemble the original sequences and analyze the large amount of data that finally answer the research questions.

The purpose of this work is to analyze in a general way the knowledge and advances to date over the expression and of the genes related with the capsaicinoid biosynthesis pathway in different species and varieties of the *Capsicum* genus in order to determine which are the most important physiological factors for the accumulation of these metabolites and thus regulate the pungency in these commercial varieties.

2.2. GENE EXPRESSION AND TRANSCRIPTOME ANALYSIS

Transcription is a process that produces an identical RNA strand in sequence with a DNA strand (gene). It is the first step in the process of gene expression and the step in which it is frequently regulated. In some cases, RNA is used for the direct production of a polypeptide (messenger RNA or mRNA). In other cases, as in the genes of ribosomal RNA (rRNA) or transfer RNA (tRNA), the transcribed RNA is the final functional product (Krebs *et al.*, 2018). The transcriptome is the set of transcripts that expresses a cell, a tissue, an organ at a certain time and under certain circumstances (Soares and Valcárcel, 2006). The identification of the complete set of transcripts, including small and large RNA molecules, new transcripts of non-annotated genes, isoforms, and gene fusion transcripts, which support the comprehensive study of transcriptome. Sequencing technologies of the next generation (NSG) allowed to sequence whole transcriptomes or RNA-Seq. This last can be defined as a set of methodologies that make use of sequencing technologies to obtain information from a set of RNA molecules from a biological sample to determine the primary sequence and the relative abundance of each RNA (Liu *et al.*, 2013).

The depth of these technologies allows to detect and sequence even transcripts with low expression. In contrast to other technologies such as microarrays, RNA-Seq achieves a high

resolution and a larger dynamic range for expression levels, and allows *de novo* annotation (Martin and Wang, 2011).

NGS provide parallel mass analysis, high performance of multiple samples and are cheaper than previous technologies such as Sanger. These technologies require the cutting of DNA, cDNA or RNA molecules into small fragments, which are subsequently sequenced. A library is created by ligating adapters, tags, primers at the ends of the fragments. Amplification of the library is done on solid surfaces or beads where there are droplets or emulsions where sequencing occurs. The incorporation of the nucleotide is carried out directly by luminescence or by changes in an electric charge during the sequencing procedure (Kulski, 2016). The small fragments of sequences, called *reads*, must be united to reconstruct the original transcriptome, a process that is called assembly. Transcriptome assembly strategies can be by comparison, in which a transcriptome or reference genome is used; and *de novo* assembly, where a reference sequence is not used, taken advantage of the redundancy of the sequencing of short fragments to find superpositions between *reads* that assemble them in transcripts (Martin & Wang, 2011).

2.3. STUDIES OF GENE EXPRESSION AND TRANSCRIPTOMICS RELATED TO CAPSAICINOIDS IN THE GENUS *Capsicum*

In *Capsicum*, several studies of gene expression have been carried out. The studies related to the biosynthesis of capsaicinoides involves the pathways of phenylpropanoids, biosynthesis of branched-chain fatty acids, and biosynthesis of precursors like phenylalanine and branched-chain aminoacids (Figure 2.1).

Differences in pungency depends in part on the genotype. The absence of capsaicinoides in certain species o varieties on *Capsicum* genus is related to mutations in genes of transcription factors and structural genes of the capsaicinoid pathway and the related ones (Arce-Rodríguez and Ochoa-Alejo, 2019). Most of these studies are related to the genes pAMT (Tsurumaki and Sasanuma, 2019; Tanaka *et al.*, 2018; Park *et al.*, 2015a) and AT3/PUN1/CS (Egan *et al.*, 2019; Stellari *et al.*, 2010; Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005) which encodes the enzymes aminotransferase (EC 2.6.1-) (Weber *et al.*, 2014; Curry *et al.*, 1999) and capsaicin synthase (Ogawa *et al.*, 2015), respectively, that catalyzes the last two steps in the synthesis of capsaicinoids. In addition, mutations which are associated with lack of pungency and capsaicinoids in *Capsicum* varieties were found in a gene of the branched-chain fatty acid biosynthesis pathway, CaKR1 (ketoacyl-ACP reductase, EC 1.1.1.100) (Koeda *et al.*, 2019) and

a gene of the transcription factor CaMYB31 (Han *et al.*, 2019). The grade of pungency also can be modulated by genes through Quantitative Trait Loci (QTL) (Park *et al.*, 2019).

In *Capsicum* fruits the tissue of the placenta is highly pungent and is the mainly site where the capsaicinoids are synthesized (Ananthan *et al.*, 2018). Studies in tissue culture of placenta in *C. chinense* Jacq. 'Umorok' showed that its content of capsaicin and dihydrocapsaicin decreases in placenta explants cultured for callus formation. During subcultures in cycles of 32 days, the main decrease was observed during the first two cycles. A comparative expression analysis between placenta tissues and callus cultures was carried out by RT-qPCR to determine the behavior of the genes involved in capsaicinoid biosynthesis. Two putative genes (*pAMT* and *PUN1*) had a similar pattern with the evolution of capsaicinoids, decreasing the first two months and greatly reduced in further cycles (Kabita *et al.*, 2019). These genes generally have shown high consistency with the content of capsaicinoids and they catalyze the last two steps in capsaicinoid pathway, which indicates that probably these are key steps.

In most plants, the localization of the capsaicinoid synthesis and accumulation has been founded in epidermal cells of the interocular septum of placenta (Tyler *et al.*, 2016; Stewart Jr *et al.*, 2007; Zamski *et al.*, 1987). An interesting phenotypic relation between capsaicinoid synthesis, the gene *AT3* and the formation of structures of accumulation of capsaicinoids (blisters) in the interocular septum has been observed (Stewart Jr *et al.*, 2007). These structures were observed during capsaicinoid accumulation in the *C. chinense* pungent variety 'Habanero' and were absent in the non-pungent variety NMCA30036, which had a recessive allele (*pun1²*) that presented an elimination of four base pairs that resulted in a reading frame mutation leading to a truncated non-functional At3 protein. In crossing experiments between the pungent and the non-pungent varieties, genotypes *pun1²/pun1²* resulted in the absence of pungency, capsaicinoids and blisters. Pungency and blisters were dominant traits (Stewart Jr *et al.*, 2007).

Another recessive allele related with the absence of capsaicinoid observed in the *C. annuum* varieties 'Jupiter', 'Maor' and Jalapeño 'Sweet 3575' was *pun1*, that presented a 2.5 kb deletion spanning the promoter and the first exon of the gene. This mutation resulted in no transcription of *AT3* (Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005). The genotype *pun1/pun1* resulted in absence of capsaicinoids and non-pungent varieties (Stewart Jr *et al.*, 2005).

In other work, another recessive allele of the gene *AT3* and named *pun1*³ was found in *C. frutescens* variety PI594141. The allele had deletions and insertions truncating the second exon and resulting in a frameshift mutation. The expression of the gene wasn't observed (Stellari *et al.*, 2010).

More recently has been observed that the gene *AT3* has a high level of conservation throughout the Solanaceae family and it present a tandem duplication: *AT3-1* and *AT3-2*. The paralog *AT3-2* is a pseudogene, while *AT3-1* is associated with pungency. Also, in certain non-pungent species such as *C. rhomboidium* there may has been in the past recombinations between *AT3-1* gene and *AT3-2* that led to the modification of the exon 2 in *AT3-1* which includes its active site and, together with mutation events in the reading frame, to the consequent loss of its functionality (Egan *et al.*, 2019).

It has been observed that the *AT3* gene is specific for fruits and closely follows the At3 protein and capsaicinoid accumulation, first increases in early or medium development stages and gradually decreases. In matures fruits and in late development stages the expression is small or negligible (Arce-Rodríguez and Ochoa-Alejo, 2017; Tanaka *et al.*, 2017; Zhang *et al.*, 2016; Arce-Rodríguez and Ochoa-Alejo, 2015; Martínez-López *et al.*, 2014; Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005).

The gene *AT3* also has an influence over the expression of other genes of the capsaicinoid pathway and the silencing of this gene reduced the capsaicinoid content and the expression of others genes (Arce-Rodríguez and Ochoa-Alejo, 2015; Stewart Jr *et al.*, 2005). In a study of the expression of genes of the capsaicinoid pathway in *C. annuum* L. cv 'Tampiqueño' a positive correlation was observed in the expression of the *AT3* gene and capsaicinoids in placentas of fruits. A pattern of increased expression was found from 10 to 40 days post-anthesis (DPA), and then decreased in the following stages (50 and 60 DPA). Interestingly, in plants in which this gene was silenced, a negative effect on the expression of *pAMT*, *BCAT* (branched-chain aminoacid transferase, EC 2.6.1.42), *KAS* (ketoacyl-ACP synthase EC 2.3.1.41) and *ACL* (acyl carrier protein), was also observed in comparison with non-silenced plants (Arce-Rodríguez and Ochoa-Alejo, 2015).

Other loci determine the presence or absence of pungency in *Capsicum*. A non-pungent variety, *C. chaconse* PI260433 had another locus named *PUN2* whose allele was responsible for the non-pungency of its fruits (Stellari *et al.*, 2010). More recently, a locus named *PUN3* that

controlled pungency in *C. annuum* was founded. The gene associated was homologous with the transcription factor CaYMB31. It was expressed in the pungent variety *Capsicum annuum* 'Teañ'. The non-pungent *C. annuum* 'YCM334' had a recessive nonfunctional allele with a nonsense mutation in the first exon and a premature stop codon leading to low expression of the gene (K. Han *et al.*, 2019). CaYMB31 controls the expression of structural genes of the capsaicinoid pathway (Han *et al.*, 2019; Arce-Rodríguez and Ochoa-Alejo, 2017).

Another gene with qualitative influence over pungency is *pAMT*. In various works the expression of this gene shows a positive correlation with capsaicinoid accumulation during the development of the fruits, similar to *AT3* (Arce-Rodríguez and Ochoa-Alejo, 2017; Tanaka *et al.*, 2017; Park *et al.*, 2015a; Martínez-López *et al.*, 2014; Stewart Jr *et al.*, 2005). There has been observed the presence of capsinoids instead of capsaicinoids in fruits where mutations in the *pAMT* gene. Capsinoids are analogs of capsaicinoids with lower pungency present in *Capsicum* genus (Antonio *et al.*, 2018). In *C. frutescens* 'S3212' a low level of capsaicinoids and high level of capsiate in comparison of pungent cultivars was observed. This variety had an allele of gene *pAMT* with a deletion in exon 7 and a SNP in exon 14 (*pamt8*), leading to a low transcription level of the gene in fruit placentas in comparison with the pungent plants (Park *et al.*, 2015a).

Other works have founded non-functional alleles in *C. chinense* varieties. 'No 4034' was a variety with low levels of capsaicinoids and high levels of capsinoids in comparison with the pungent 'Red Habanero'. The recessive allele in 'No 4034' (*pamt9*) had a 7 pb insertion in the exon 16 that resulted in a frameshift mutation leading to low expression of the gene in comparison with 'Red Habanero' (Tanaka *et al.*, 2018). In addition, it has been observed that most of the mutations in *pAMT* alleles in *C. chinense* varieties are insertions of transposons (Tanaka *et al.*, 2018; 2015; 2010; Koeda *et al.*, 2014).

In *C. annuum* 'Color Piman Yellow' there was found an allele of *pAMT* (*pamt¹⁰*) with a SNP at the exon 11 that resulted in a premature stop codon and a truncated protein sequence, leading to a low expression of the gene in comparison with the pungent *C. annuum* 'Takanotsume' that had a non-mutated gene. The non-pungency was attributed to the *pamt* allele (Tsurumaki and Sasanuma, 2019).

In the non-pungency pepper *C. chinense* 'No. 3341' a single recessive allele of CaKR1 had an insertion of a transposable element of 4.5 kb in the first intron, resulting in the production of a truncated protein missing the region of the catalytic domain and it caused the lack of

capsaicinoids, 8-methyl-6-nonenic acid and pungency. In contrast, in pungents *C. chinense* 'Habanero pepper' and *C. annuum* 'NMH' the gene was expressed. When the gene was silenced in 'NMH' plants, there was a significant reduction of its expression and in the content of capsaicinoids of (Koeda *et al.*, 2019). This enzyme is part of the pathway of branched-chain fatty acid biosynthesis for the formation of the 8-methyl-6-nonenyl-CoA, which condenses with vanillylamine for the formation of capsaicin (Figure 2.1).

Others studies compared the gene expression among different tissues, organs, development stages and conditions in plants. This is useful for identify new genes and predict its biologic role and to study the influence of environmental factors in the expression of genes (Arce-Rodríguez and Ochoa-Alejo, 2019). A common strategy to identify new genes and to investigate its role in the capsaicinoid pathway is to compare pungent and non-pungent tissues. A study related to the expression of genes of the capsaicinoid pathway in different parts of the fruit (seeds, pericarp and placenta) and in different varieties of *C. chinense* found that certain varieties show synthesis of an expression of genes of that pathway for placenta and pericarp (Tanaka *et al.*, 2017).

In most varieties of plants of genus *Capsicum*, the capsaicinoids are synthesized in placental tissues. In pericarp there has been observed the existence of capsaicinoids, vesicles and elongated epidermal cells secreting capsaicinoids, similar to those observed in the placental septum (Sugiyama, 2017; Bosland *et al.*, 2015). In the work of (Tanaka *et al.*, 2017) the varieties studied were *C. chinense* 'Trinidad Moruga Scorpion Yellow' (MY) and 'Red Habanero' (HB), *C. annuum* 'Takanotsume' (TK) and the non-pungent variety 'Fushimi-amanaga' (FA). The analysis found that MY and HB had the highest capsaicinoid content, and the highest levels were detected in the placental septum and in the pericarp of MY. Also, a qRT-PCR analysis found a strong expression of *PUN1*, *pAMT*, *KAS* and *BCAT* in placental tissues of both varieties and in the pericarp of MY. The expression of these genes was positively associated with the production of capsaicinoids in the pungent tissues: low expression levels at the beginning of fruit development and peaked at 30 days after flowering (DAF), ripe green fruit stage. In the stage of fully mature fruit (40 DAF), the expression of genes *PUN1*, *pAMT* and *KAS* decreased rapidly in placentas of both varieties and in the pericarp of MY. In contrast, *BCAT* only decreased in placentas and peaked at 40 DAP in the pericarp for both varieties (Tanaka *et al.*, 2017). The *KAS* and *BCAT* genes provide branched-chain fatty acids precursors for capsaicinoid formation. The latter is also an essential gene in the catabolism of branched-chain amino acids. These two roles can explain its different behavior (Tanaka *et al.*, 2017; Curry *et al.*, 1999).

Another research studied by qRT-PCR the participation of a transcription factor CaMYB31 in the varieties of *C. annuum* 'Tampiqueño 74' (pungent) and 'California Wonder' (non-pungent) (Arce-Rodríguez and Ochoa-Alejo, 2017). It had an expression positively associated with the synthesis of capsaicinoids in placental tissues of fruits for the variety 'Tampiqueño 74'. The content of capsaicin and dihydrocapsaicin increased during the fruit development and had a peak at 40 DPA. The expression of the *CaMYB31* gene was positively correlated with the expression of the genes *C4H* (cinnamate 4-hydroxylase, EC 1.14.14.91), *COMT* (caffeoyl-CoA O-methyltransferase, EC 2.1.1.104), *KAS*, *pAMT* and *AT3*: they increased during the fruit development, with a maximum at 30 or 40 DPA and then decreased. The expression of these genes, except *COMT* and *C4H*, was practically not observed in placental tissues of 'California Wonder' (Arce-Rodríguez and Ochoa-Alejo, 2017). In addition, the genetic silencing of *CaMYB31* significantly reduced the expression of various genes, including *KAS*, *BCAT*, *C4H*, *COMT*, *pAMT* and *ACL*, as well as the production of capsaicinoids. It has been proposed that *CaMYB31* may act as a master regulator of multiple genes of the capsaicinoid pathway (Han *et al.*, 2019; Arce-Rodríguez and Ochoa-Alejo, 2017).

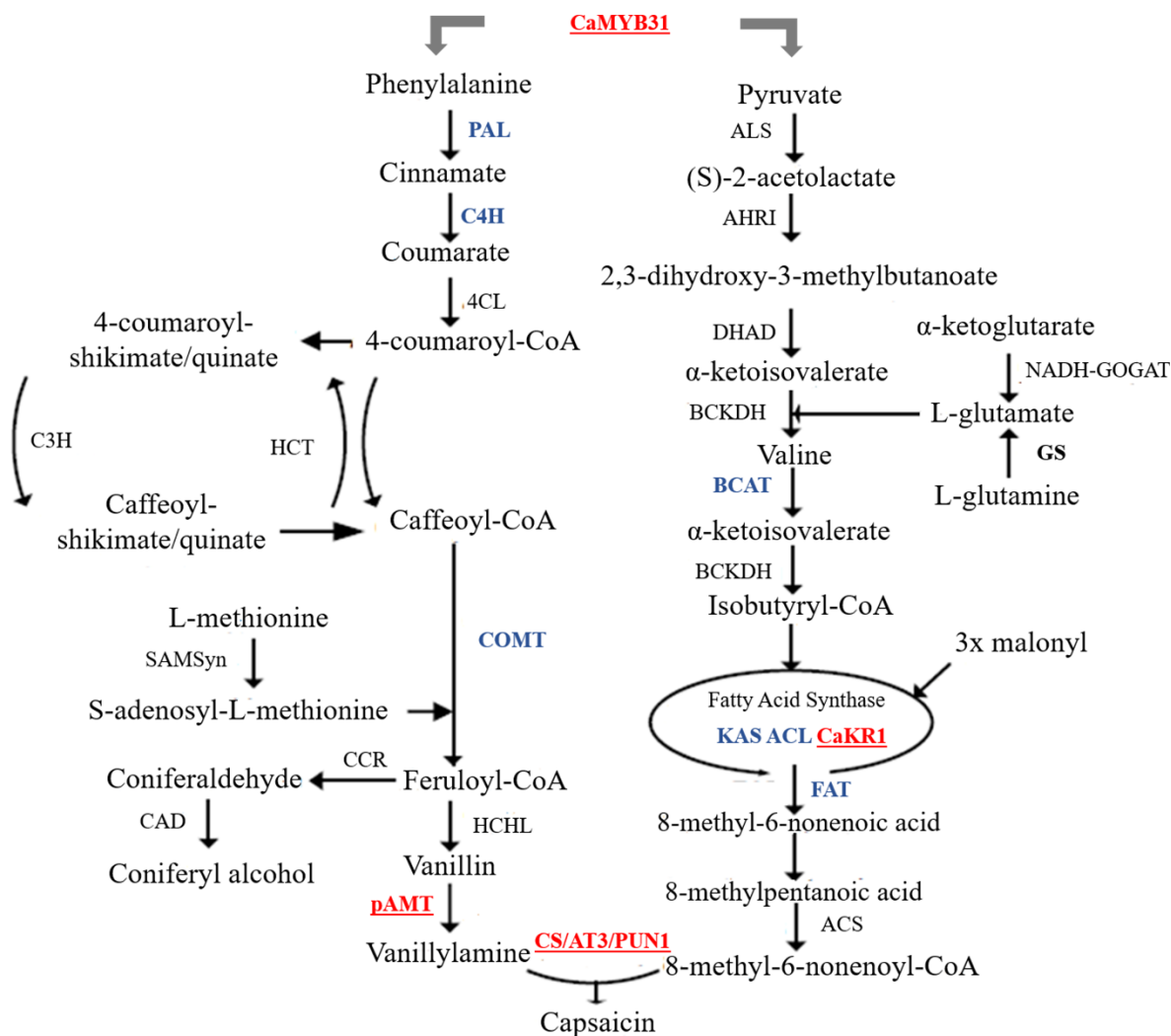


Figure 2.1. Capsaicinoid biosynthesis model. The enzymes, whose alleles determine the pungency or non-pungency of the plants, are in red and underlined. Other enzymes, that have consistency or association with changes in the synthesis of capsaicinoids, are in blue. CaMYB31 is a transcription factor that has an influence on the expression of genes in the phenylpropanoids and branched-chain fatty acid pathways. PAL: phenylalanine ammonia lyase EC 4.3.1.24; C4H: cinnamate-4-hydroxylase EC 1.14.14.91; 4CL: 4-coumaroyl-CoA ligase EC 6.2.1.12; CoA: coenzyme A; HCT: hydroxycinnamoyl transferase EC 2.3.1.133; C3H: coumaroyl shikimate / quinate 3-hydroxylase EC 1.14.14.96; COMT: caffeoyl-CoA O-methyltransferase 2.1.1.104; SAMSyn: S-adenosylmethionine synthetase EC 2.5.1.6; HCHL: hydroxycinnamoyl-CoA hydratase / lyase EC 4.2.1.101, EC 4.1.2.41; CCR: cinnamoyl-CoA reductase EC 1.2.1.44; CAD: cinnamoyl alcohol dehydrogenase EC 1.1.1.195; pAMT: putative aminotransferase EC 2.6.1.-; ALS: acetolactate α synthase EC 2.2.1.6; AHRI: acetohydroxy acid reductoisomerase EC 1.1.1.86; DHAD: dihydroxy acid dehydratase EC 4.2.1.9; GS: glutamine synthetase EC 6.3.1.2; NADH-GOGAT: nicotinamide adenine dinucleotide-dependent glutamate synthase EC 1.4.1.14; BCAT: branched chain amino acid transferase EC 2.6.1.42; BCKDH: branched-chain α -ketoacid dehydrogenase/decarboxylase EC 1.2.1.25; KAS: ketoacyl-ACP synthase EC 2.3.1.41; CaKR1: ketoacyl-ACP reductase, EC 1.1.1.100; ACP: conveyor of acyl groups; ACL: acyl group transport protein; FAT: acyl-ACP thioesterase EC 3.1.2.14; ACS: acyl-CoA

synthetase EC 6.2.1.3; CS/PUN1: capsaicinoid synthase; AT3: acyltransferase. Adapted from Arce-Rodríguez and Ochoa-Alejo (2019)

Environment factors often affect the content of capsaicinoids. The promoter of *PUN1* have motifs and consensus elements sequences that respond to light, heat stress and mechanical stress (Kim *et al.*, 2009). It can be induced by factors such wounding treatments, heat and the capsaicinoid biosynthesis pathway precursors phenylalanine and valine. Also, its expression is different between tissues in the same plant. Constructs of this promoter with the genes *GUS* (β -glucuronidase) and *NtKED* (a tobacco wound-responsive gene) were expressed in tobacco and tomato, respectively (Kirke *et al.*, 2018). In tobacco plants, the increase in the expression of *GUS* by phenylalanine was stronger than the observed by valine. The heat treatment (40 °C) and mechanical stress induced a significant expression of *GUS* in leaves of tobacco in comparison to untreated plants. In addition, the promoter directed expression of *GUS* in most tissues of tobacco, but it was stronger in buds, young anther, pollen and others reproductive tissues than in root or leaf. In tomato leaves, the promoter responded to wounding stress but heat treatment did not induce a significant increase in the transcription of *NtKED*. Also, the expression of *NtKED* with this promoter was stronger in red and green fruits than in petals or leaves (Kirke *et al.*, 2018). In *Arabidopsis thaliana* with construct of the promoter and *GUS* the light induced the expression more than dark or dim light conditions. The expression of *GUS* was stronger at heat stress of 37 °C than plants incubated at 23 °C. In cutting edges of stems (mechanical stress) in transgenic plants there was observed more expression of *GUS* than in plants with empty vector. Also wounding stress stimulated the expression of *CS* in placentas of green fruits in *C. annuum* (Kim *et al.*, 2009). It was observed that capsaicin reduces the expression of *CS*, *KAS*, *pAMT* and *PAL* (phenylalanine ammonia-lyase, EC 4.3.1.24). This negative feedback is a possible reason for which the peak of expression of genes like *CS* occurs before the peak of capsaicinoids during fruit development (Kim *et al.*, 2009).

The expression of *CaMYB31*, *KAS* and *pAMT* has been affected in fruits of *C. annuum* by hormones such as indoleacetic, jasmonic, salicylic and gibberellic acids or by mechanical stress, light and temperature (Arce-Rodríguez and Ochoa-Alejo, 2017). With the application of salicylic acid, an increase gene expression was observed, while with jasmonic acid a decrease was observed. Wound stress showed an effect in decreasing gene expression. In the evaluation of temperature (4 °C, 25 °C and 37 °C), higher expression was observed at the lower temperature. In light treatment, fruits exposed to fluorescent light presented stronger expression than those exposed to dark (Arce-Rodríguez and Ochoa-Alejo, 2017). It has been found that

salicylic acid and methyl jasmonate induce the production of capsaicinoids, intermediates or the activity of enzymes of the biosynthetic pathway, but it also depends on concentration and time of exposure (Rodas-Junco *et al.*, 2013; Altúzar-Molina *et al.*, 2011; Gutiérrez-Carbajal *et al.*, 2010).

The temperature also has an effect on the content of capsaicinoids in fruits, being able to increase or decrease according to the temperature of growth and is different between varieties (González-Zamora *et al.*, 2013). On the other hand, the content of capsaicinoids is different in the fruits of plants exposed to light of different wavelengths or fluorescent lights (Gangadhar *et al.*, 2012).

Water management is another factor that affect the capsaicinoid biosynthesis. In *C. chinense* there was a tendency of increase the capsaicinoid content in placental tissues of fruits of plants treated by withholding irrigation for 7 or 9 days in comparison with plants irrigated each day. Also, water stress induced the activity of CS and decrease the activity of peroxidases (POD) in comparison with the control during fruit development (Zamudio-Moreno *et al.*, 2014). In another work, it was found that under drought stress the content of capsaicin and the activity of the enzymes PAL, C4H, CS and POD increases in comparison with non-stressed *Capsicum* plants (Phimchan *et al.*, 2014). The changes in transcription levels of the genes of the enzymes was not investigated in water stress and requires further attention.

Studies by RNA-Seq allows the analysis of thousands of sequences of a transcriptome. Han *et al.* (2019) realized a RNA-Seq differential analysis between placental tissues in *C. annuum* 'Tean' and 'YCM334'. The results revealed that several structural genes of the capsaicinoid pathway (*PUN1*, *KAS*, *BCAT* and *FAT* (acyl-ACP thioesterase, EC 3.1.2.14)) had low expression in the non-pungent variety YCM334 in comparison with the pungent variety Tean in fruits between 26 and 36 DPA (Han *et al.*, 2019).

Another study in placenta and pericarp of *C. frutescens* 'Xiaomila' used RNA-Seq analysis and *de novo* assembly to analyze transcripts for genes of the biosynthesis routes related to the synthesis of capsaicinoids expressed between 20 and 30 DAF (Liu *et al.*, 2013). The identified sequences had predicted functions for the metabolic pathways of phenylpropanoids and benzenoids, branched-chain fatty acid biosynthesis, phenylalanine metabolism and branched-chain amino acid biosynthesis, finding 197 transcripts for putative sequences of these routes, including 35 known and 162 new candidates from previous metabolic pathways. In addition,

they found sequences not previously reported in the pathway, threonine deaminase (TD, EC 4.3.1.19), dihydroxyacid dehydratase (DHAD, EC 4.2.1.9) and prephenate aminotransferase (PAT, EC 2.6.1.79). The first two participate in the branched-chain amino acid biosynthesis pathway and the last belongs to the transaminases (Liu *et al.*, 2013).

Another study analyzed the change on the expression of genes from the capsaicinoid biosynthetic pathway in placental tissues of *C. frutescens* 'Guijiangwang' by RNA-Seq during five stages of fruit development: 10, 20, 30, 40 and 50 DAP. There were 2340 genes differentially expressed in consecutive stages and 131 genes were annotated as related to the pathway of capsaicinoid synthesis. Among the genes that followed an expression pattern consistent with the accumulation of capsaicinoids *PAL*, *C4H*, *ACS* (acyl-CoA synthetase, EC 6.2.1.3), *NADH-GOGAT* (nicotinamide adenine dinucleotide-dependent glutamate synthase, EC 1.4.1.14), *AT3*, *KAS*, *COMT* and *BCKDH* (branched-chain α -ketoacid dehydrogenase/decarboxylase, EC 1.2.4.4) were included (Zhang *et al.*, 2016).

Other wide transcriptome study was realized in whole fruits of *C. annuum* 'Tampiqueño' at 10, 20, 40 and 60 DPA. By *de novo* assembly, they obtained 34066 genes that classified according to their expression patterns between consecutive periods. The most active period of expression changes was from 40 to 60 DPA, characterized by a general decrease in the level of gene expression (Martínez-López *et al.*, 2014). Regarding the biosynthesis of capsaicinoids, the average expression pattern of the genes of this pathway was increase-decrement-decrement (IDD), indicating a peak at 20 DPA. Several individual genes of the route such as those corresponding to *PUN1*, *BCAT*, *COMT*, *PAL*, *FAT*, *pAMT* and *KAS* had this behavior, while others such as *ACS* had a pattern of continuous expression increase throughout the development of the fruit (Martínez-López *et al.*, 2014).

Other work reported a comparative analysis of the expression of genes related with the capsaicinoid biosynthesis in placental tissues of *C. annuum* 'CM334' and a non-pungent variety 'ECW' in different fruit development stages. The orthologues in tomato were also compared. The expression of *CS* was detected exclusively in placentas of the pungent variety. Also, *BCAT* and *KAS* were not greatly expressed in tomato in the same development period. Between pungent and non-pungent plants *BCAT*, *COMT* and *FAT* showed a differential expression at six DPA and *CS* was barely expressed in 'ECW'. These results suggest the importance of these genes and its expression patterns for capsaicinoid synthesis in pungent plants (Kim *et al.*, 2014).

Regarding QTL analysis for the content of capsaicinoids in *Capsicum*, in a study the segregation of three capsaicinoids (capsaicin, dihydrocapsaicin and nordihydrocapsaicin) was analyzed in an interspecific cross between a mildly pungent *C. annuum* and a very pungent *C. frutescens*. Six QTL that controlled the content of capsaicinoids were detected on chromosomes 3, 4 and 7. Five were for capsaicin, four for dihydrocapsaicin and one for nordihydrocapsaicin. Most of the QTL were overlapping, with the exception of one for capsaicin and the only QTL found for nordihydrocapsaicin. The EST (expressed sequence tag) 3A2 and the gene *BCAT* were localized in QTL detected on chromosomes 3 and 4, respectively. 3A2 apparently is related with valine catabolism (Ben-Chaim *et al.*, 2006).

Park *et al.* (2019) did an analysis of QTL in the pericarp of *Capsicum* fruits that were related to the biosynthesis of capsaicinoids. Using QTL-Seq in crosses of the varieties *C. chinense* 'Habanero' (with pungent placenta and non-pungent pericarp) and 'Jolokia' (pungency in pericarp and placenta) they found a QTL (HJ16-qt1seq) on chromosome 6 that seemed to be related to the control of the content of capsaicinoids in the pericarp. Also, a QTL mapping analysis in crosses of 'Habanero' and 'Jolokia' and 'Jolokia' with *C. chinense* 'SNU11-001' (pungent placenta and non-pungent pericarp) found several QTL associated with capsaicin, dihydrocapsaicin and total capsaicinoids. Some of these were commonly localized on the chromosome 6 in the two mapping populations and they were close to the HJ16-qt1seq region. A comparative analysis by RNA-Seq carried out in this same work in the varieties 'Habanero' and 'Scorpion' (pungent in pericarp and placenta) revealed that several genes of the metabolism of capsaicinoid biosynthesis were highly expressed in the pericarp of 'Scorpion' compared with the pericarp of 'Habanero', including *PUN1*, *BCAT*, *KAS*, *4CL* (4-coumaroyl-CoA ligase, EC 6.2.1.12), *C4H*, *KR* (*CaKR*) and *FAT*. The last three genes were located in (or very close to) the QTL on chromosome 6 associated with capsaicinoids (Park *et al.*, 2019).

In the Table 2.1 there was a resume with the principal points of works reviewed here related with genes and molecular studies of the capsaicinoid biosynthesis pathway. In Table 2.2 there was a resume of the principal findings in the transcriptome studies reviewed here. Also, Figure 2.1 shows the capsaicinoid biosynthesis pathway with its enzymes. The enzymes whose genes have presented alleles that determine the pungency or non-pungency of the plants are in bold red and underlined. Other enzymes that in various works have showed consistency or association with changes in the synthesis of capsaicinoids are in blue bold and in italics.

Table 2.1. Studies related to genes and molecular findings of capsaicinoid synthesis pathway

Author	Principal findings
Stewart Jr <i>et al.</i> , 2005	A recessive allele of <i>PUN1</i> (<i>pun1</i>) is related with the absence of pungency in varieties of <i>C. annuum</i>
Ben-Chaim <i>et al.</i> , 2006	In <i>C. annuum</i> the segregation of capsaicin, dihydrocapsaicin and nordihydrocapsaicin was analyzed. Six QTL controlling capsaicinoid content were founded on chromosomes 3, 4 and 7. Most of the QTL were on chromosomes 4 and 7. The gene <i>BCAT</i> was localized on chromosome 4. An EST related with valine catabolism was localized on chromosome 3
Stewart Jr <i>et al.</i> , 2007	A recessive allele of <i>PUN1</i> (<i>pun1²</i>) contribute to non-pungency of varieties of <i>C. chinense</i> . The allele was cosegregated with the lack of capsaicinoids and blisters
Kim <i>et al.</i> , 2009	The promoter of <i>PUN1</i> has have motifs and consensus elements that respond to light, heat stress and mechanical stress. The light, mechanical stress and heat stress induced the promoter. The capsaicin had negative feedback over the expression of structural genes of the capsaicinoid pathway.
Stellari <i>et al.</i> , 2010	A recessive allele of <i>PUN1</i> (<i>pun³</i>) is related with the absence of pungency in varieties of <i>C. frutescens</i> . In <i>C. chacoense</i> the lack of pungency is related with the locus <i>PUN2</i> , which is not allelic to <i>PUN1</i> .
Tanaka <i>et al.</i> , 2010	Non-functional alleles of <i>pAMT</i> (<i>pamt³</i> , <i>pamt⁴</i> , <i>pamt⁵</i>) were founded in mildly-pungent <i>C. chinense</i> varieties which present high levels of capsinoids and low level of capsaicinoids. Some mutations in <i>C. chinense</i> are insertions of transposons.
Koeda <i>et al.</i> , 2014	A non-pungent variety of <i>C. chinense</i> presented a non-functional allele of <i>pAMT</i> (<i>pamt⁶</i>).
Arce-Rodríguez and Ochoa-Alejo, 2015	The gene <i>AT3</i> had influence in the expression of other genes of the capsaicinoid biosynthesis pathway. Silencing this gene reduced the expression of <i>pAMT</i> , <i>BCAT</i> , <i>KAS</i> and <i>ACL</i>
Park <i>et al.</i> , 2015a	In varieties of <i>C. frutescens</i> an allele of <i>pAMT</i> (<i>pamt⁸</i>) leaded to a low level of capsaicinoids in fruits. Capsinoids were in high level.
Tanaka <i>et al.</i> , 2015	A low-pungent variety of <i>C. chinense</i> which preferentially produced capsinoids over capsaicinoids had a non-functional allele of <i>pAMT</i> (<i>pamt⁷</i>) whose sequence contained a mutation by transposon insertion.
Arce-Rodríguez and Ochoa-Alejo, 2017	The transcription factor <i>CaYMB31</i> controls the expression of structural genes of the capsaicinoid pathway. In pungent varieties of <i>C. annuum</i> the expression of the <i>CaMYB31</i> gene was positively correlated with the expression of <i>C4H</i> , <i>COMT</i> , <i>KAS</i> , <i>pAMT</i> and <i>AT3</i> . The genetic silencing of this gene reduced the expression of structural genes of the pathway and the content of capsaicinoids. The expression of <i>CaMYB31</i> , <i>KAS</i> and <i>pAMT</i> was affected in fruits by indoleacetic, jasmonic, salicylic and gibberellic acids, mechanical stress, light and temperature
Tanaka <i>et al.</i> , 2017	Varieties of <i>C. chinense</i> present capsaicinoids in pericarp and placenta and strong expression of genes related with this pathway (<i>PUN1</i> , <i>pAMT</i> , <i>KAS</i> and <i>BCAT</i>). Expression was positively associated with capsaicinoid production.
Kirke <i>et al.</i> , 2018	The <i>PUN1</i> promoter can be induced by phenylalanine, valine, heat stress and mechanical stress. In tomato and tobacco, the expression of genes with this promoter is stronger in reproductive tissues than in vegetative tissues.
Tanaka <i>et al.</i> , 2018	An allele of <i>pAMT</i> (<i>pamt⁹</i>) was found in low-pungent varieties of <i>C. chinense</i> . The gene had a low expression in comparison with pungent varieties.
Egan <i>et al.</i> , 2019	The gene <i>AT3-1</i> and its paralog, a pseudogene called <i>AT3-2</i> , are present throughout the Solanaceae family, and they have a high level of

	conservation. In <i>C. rhomboidium</i> there has been recombinations events between <i>AT3-1</i> and <i>AT3-2</i> leading to the loss of pungency in this specie.
Han <i>et al.</i> , 2019	The locus <i>PUN3</i> has influence over the expression of varieties of <i>C. annuum</i> . This locus had a sequence homologous with the transcription factor <i>CaYMB3</i> .
Kabita <i>et al.</i> , 2019	Placental tissues were used to induce callus formation. In the culture and subcultures de content of capsaicinoids and the expression of genes involved in capsaicinoid pathway were analyzed. <i>pAMT</i> and <i>PUN1</i> had a similar pattern of expression with the evolution of capsaicinoids
Koeda <i>et al.</i> , 2019	A non-pungent variety of <i>C chinense</i> had recessive allele of the gene <i>CaKR1</i> , and it was related with the lack of capsaicinoids. When the gene was silenced in pungent plants, there was a reduction in the content of capsaicinoids
Park <i>et al.</i> , 2019	Analysis of QTL and genes related with capsaicinoid biosynthesis pathway in <i>C. chinense</i> varieties with pungent pericarps. Various QTL related with capsaicinoid synthesis were localized on chromosome 6, and the genes <i>C4H</i> , <i>KR</i> and <i>FAT</i> were assigned in it.
Tsurumaki and Sasanuma, 2019	A non-pungent variety of <i>C. annuum</i> 'had an allele of <i>pAMT</i> (<i>pamt¹⁰</i>) which leads to no production of capsaicinoids.

Table 2.2. Transcriptomic studies related with capsaicinoid synthesis pathway

Author	Principal findings
Liu <i>et al.</i> , 2013	In <i>C. frutescens</i> identification and prediction of functions of 197 transcripts for putative sequences of genes of pathways related with capsaicinoid biosynthesis, including 162 new candidates. Sequences not previously reported in the pathway were founded: <i>TD</i> , <i>DHAD</i> and <i>PAT</i>
Kim <i>et al.</i> , 2014	In <i>C. annuum</i> a comparative analysis of the transcriptome in placental tissues of pungent and non-pungent varieties was made. Some genes like <i>CS</i> , <i>BCAT</i> , <i>COMT</i> and <i>FAT</i> presented differential expression in some points of the fruit development., suggesting the importance of the genes and its expression pattern
Martínez-López <i>et al.</i> , 2014	In <i>C. annuum</i> the dynamic of wide transcriptomic of the whole fruit was analyzed during its development. The average expression pattern of the genes related with capsaicinoid pathway was increase-decrement-decrement with a peak at 20 DPA. Various structural genes of the pathway showed this patron, including <i>PUN1</i> , <i>BCAT</i> , <i>COMT</i> , <i>PAL</i> , <i>FAT</i> , <i>pAMT</i> and <i>KAS</i>
Zhang <i>et al.</i> , 2016	In placental tissues of <i>C. frutescens</i> the dynamic of the expression of genes related with the capsaicinoid synthesis was analyzed during fruit development. Various genes showed an expression pattern consistent with capsaicinoids accumulation, including <i>PAL</i> , <i>C4H</i> , <i>ACS</i> , <i>NADH-GOGAT</i> , <i>AT3</i> , <i>KAS</i> , <i>COMT</i> and <i>BCKDH</i>
Han <i>et al.</i> , 2019	Differential analysis of expression in placental tissues of pungent and non-pungent varieties <i>C. annuum</i> varieties. Various genes of the capsaicinoid pathway have a highly expression on pungent varieties compared with the non-pungent varieties, including <i>PUN1</i> , <i>KASI</i> , <i>BCAT</i> and <i>FAT</i> between 26 and 36 DPA
Park <i>et al.</i> , 2019	Comparative analysis between <i>C. chinense</i> varieties with pungent and non-pungent pericarps revealed that several genes of the capsaicinoid pathway were highly expressed in pungent pericarp, including <i>PUN1</i> , <i>BCAT</i> , <i>KAS</i> , <i>4CL</i> , <i>C4H</i> , <i>KR</i> and <i>FAT</i>

In conclusion, the different molecular and transcriptomic approaches applied to different species and varieties within the genus *Capsicum* show that factors like the different stages of development of fruits, the type of tissue (placenta or pericarp), environmental factors and abiotic stress can drastically change the expression of the transcriptomic profile of plants, and in some cases these genetic changes are related to the production of capsaicinoids. Also, the genotype of the plant can determine the pungency or non-pungency of a plant by the alleles of key genes like *AT3*, *pAMT* or *CaMYB31*, which it carried. In addition, the genetic constitution controls the capsaicinoid by set of genes which modulate the synthesis of capsaicinoids, like QTL or the gene regulation in the capsaicinoid biosynthesis pathway and also other related. The transcriptomic analyzes applied indicate that in addition to the traditional pathway of capsaicinoid biosynthesis, other related pathways are involved, such as the synthesis of phenylpropanoids and the biosynthesis of precursor amino acids. Finally, the variety of the plants and the specific tissue of analysis, like whole fruit, placenta or pericarp, are factors that should also be considered when evaluating the transcriptomic profile of these plants, their relationship with the accumulation of capsaicinoids and the highly appreciated pungency characteristic of these fruits. All of this knowledge is valuable in plant breeding to control the pungency of the crops. Further studies are necessary to deepen more in the understanding of the transcriptome profile related with the capsaicinoid biosynthesis, its genetic control, and its relation with other pathways and its interaction with the environment, especially with those key factors that affect the plant development, production and productivity in crops.

CAPÍTULO III**EXPRESIÓN DIFERENCIAL DE GENES RELACIONADOS CON EL DESARROLLO Y DE SÍNTESIS DE CAPSAICINOIDES EN FRUTOS DE PLANTAS DE CHILE HABANERO (*Capsicum chinense* Jacq.) CRECIDAS EN DIFERENTES TIPOS DE SUELO****RESUMEN**

El chile habanero (*Capsicum chinense* Jacq.) es una planta de importancia económica en la Península de Yucatán (México). Esta planta destaca por su alto contenido en capsaicinoides en relación a chiles de otras regiones del mundo y que le dan la sensación de picor a los frutos.

El suelo es uno de los factores que más influyen en el desarrollo de los frutos, su calidad nutritiva y el contenido de capsaicinoides. Sin embargo, el efecto del tipo de suelo en el desarrollo de los frutos y en el metabolismo de capsaicinoides en cultivos de chile habanero apenas ha sido estudiado.

El presente trabajo tuvo como objetivo comparar el efecto de suelos con características contrastantes, suelo negro (SN) y rojo (SR), sobre la expresión de genes relacionados con el desarrollo general y la síntesis de capsaicinoides a través del análisis transcriptómico de frutos de chile habanero.

Las plantas crecidas en SR tuvieron frutos de 45 DPA con un mayor tamaño y una alta expresión de genes relacionados con el desarrollo floral, el suavizamiento y la abscisión, lo que indica un favorecimiento de desarrollo de frutos desde las etapas tempranas hasta la madurez.

En el SN se enriquecieron rutas metabólicas para el crecimiento, azúcares y fotosíntesis. También favoreció la acumulación de capsaicinoides, alcanzando un máximo a 25 DPA, antes que SR. Se activaron algunos genes posiblemente relacionados con la síntesis de valina y leucina (*KARI*), la actividad antioxidante de los frutos y producción de NADPH (*G6PDH*), el transporte y almacenamiento vesicular de capsaicinoides en las células (*RPT4*).

Palabras clave: *Capsicum chinense*, suelo, capsaicinoides, expresión de genes

3.1. INTRODUCCIÓN

Las plantas del género *Capsicum* son un grupo que pertenece a la familia de las Solanáceas junto a otras especies como el tomate y el tabaco, y se caracterizan por tener frutos picantes. En este grupo se encuentran especies domesticadas como *C. annuum*, *C. chinense*, *C. baccatum*, *C. frutescens* y *C. pubescens* (Bosland y Votava, 2012). Los frutos de estas plantas tienen amplia importancia comercial y son consumidos tanto en fresco como en condimentos. Producen una gran variedad de metabolitos de interés industrial y médico (Akhtar *et al.*, 2021; Antonio *et al.*, 2018). Se usan en la elaboración de una gran variedad de productos, como son: fármacos, cosméticos, pigmentos, colorantes, pinturas, oleorresinas, antimicrobianos, entre otros (Akhtar *et al.*, 2021).

El chile habanero (*C. chinense* Jacq.) es una planta económica y culturalmente relevante en la Península de Yucatán en el sureste de México (Ruiz-Lau *et al.*, 2011), donde las condiciones ambientales han permitido desarrollar variedades con propiedades particulares que les han dado una Denominación de Origen (González-Estrada *et al.*, 2018). Las condiciones ambientales en las que crece las plantas afectan su desarrollo, la producción de frutos y su contenido de metabolitos (Medina-Lara *et al.*, 2019; Das *et al.*, 2016; Zamudio-Moreno *et al.*, 2014; Medina-Lara *et al.*, 2008).

En particular los suelos con características fisicoquímicas diferentes muestran afectar la producción de frutos, así como su contenido nutricional y de metabolitos (Das *et al.*, 2016). En Yucatán existen diferentes tipos de suelo, entre los que se encuentran los suelos negros (*Boxlu'um*) y los suelos rojos (*Haylu'um*), que se diferencian en propiedades fisicoquímicas como el contenido de arena, materia orgánica, nitrógeno, fósforo, carbonatos y calcio, que suelen ser mayores en los suelos negros (Estrada-Medina *et al.*, 2016; Bautista-Zúñiga *et al.*, 2003), y en sus poblaciones de bacterias y hongos (Estrada-Medina *et al.*, 2016). Se ha observado que los tallos, hojas y pedúnculos de plantas de chile habanero crecidas en suelos negro y rojo de Yucatán tienen diferente contenido de metabolitos como polifenoles, carotenos, vitamina A, vitamina C y vitamina E, y las diferencias no sólo dependen del tipo de suelo, sino también de las partes de la planta en donde se encuentran los metabolitos (Chel-Guerrero *et al.*, 2021).

Al comparar suelos negro y rojo regados con agua con alta conductividad se observó en suelo rojo un uso más eficiente de nutrientes como nitrógeno, fósforo y potasio por las plantas, y

mayor producción de frutos de mayor tamaño, mientras que en suelo negro hubo una aparente mayor retención de dichos nutrientes en suelo, pese a tener un mayor contenido, y un uso menos eficiente de éstos en la planta (Medina-Lara *et al.*, 2019).

De entre los metabolitos de chile, los capsaicinoides se caracterizan por dar a los frutos su sensación de picor, por lo que se usan en la elaboración de condimentos y repelentes, además que son interesantes para usarse en la industria, medicina y farmacéutica por sus propiedades bioactivas (Wang *et al.*, 2022). Químicamente se componen de un grupo de vanillilamina y una cadena hidrocarbonada de 9 a 11 carbonos unida a la vanillilamina a través de un grupo amida (Werner, 2021), cuya estructura da lugar a los distintos capsaicinoides, formando la capsaicina y la dihidrocapsaicina alrededor del 90% de la composición (Antonio *et al.*, 2018).

La ruta metabólica de síntesis de capsaicinoides viene de la conjunción de las rutas metabólicas de los fenilpropanoides que da lugar al grupo de la vanillilamina, y de los aminoácidos de cadena ramificada que da lugar al grupo acilo (Arce-Rodríguez y Ochoa-Alejo, 2019), convergiendo en una condensación de ambos grupos por la enzima AT3 (capsaicina sintasa o CS) (Ogawa *et al.*, 2015). Los capsaicinoides se sintetizan y acumulan en vacuolas de las células epidérmicas del septum interocular en la placenta de los frutos (Tyler *et al.*, 2016). Existe también evidencia de su acumulación en el pericarpio de frutos de variedades muy picantes en *C. chinense* (Tanaka *et al.*, 2021).

En *C. chinense* se sabe que el tipo de suelo afecta el contenido de capsaicinoides, y esto está relacionado con el contenido de materia orgánica, el nitrógeno, el fósforo, el potasio y la actividad microbiana de los suelos (Das *et al.*, 2016). En suelos negro y rojo de Yucatán, el contenido de metabolitos en los frutos de chile habanero fue mayor para las plantas crecidas en suelo rojo en comparación a los frutos de plantas crecidas en suelo negro en frutos inmaduros, de maduración intermedia y maduros (Rodríguez-Buenfil *et al.*, 2018).

El estudio de la expresión de los genes relacionados con la síntesis de capsaicinoides también es importante porque aporta información necesaria para proponer posibles estrategias que permitan controlar la producción de estos metabolitos ante determinadas condiciones ambientales. Entre los genes clave en la determinación del contenido de capsaicinoides se encuentran *AT3 (PUN1)*, *pAMT* y *MYB31* (Egan *et al.*, 2019; Han *et al.*, 2019; Tanaka *et al.*, 2019).

Se sabe que factores ambientales tales como la luz, la temperatura, el daño mecánico, la sequía, el estrés salino, la fuente de nitrógeno en el medio, así como las fitohormonas afectan la expresión de los genes de la ruta de síntesis de capsaicinoides (Zhang *et al.*, 2020a; Sun *et al.*, 2019; Kirke *et al.*, 2018; Arce-Rodríguez y Ochoa-Alejo, 2017). Sin embargo, se conoce poco sobre cómo afecta el tipo de suelo la expresión de los genes de la ruta de biosíntesis de capsaicinoides, sólo habiéndose estudiado el comportamiento de *PUN1* de manera individual (Das *et al.*, 2021), pero sobre el comportamiento de la mayoría de los genes o su patrón de expresión como un conjunto en la ruta de síntesis se desconoce, pese a la importancia que tiene el suelo sobre la producción de capsaicinoides en los frutos.

Para aumentar el conocimiento sobre el efecto del suelo sobre los genes de la ruta de síntesis de capsaicinoides y genes que pudieran estar relacionados con la producción de estos metabolitos, en este estudio se realizó una comparación del efecto que tienen los suelos negro y rojo de Yucatán sobre varios genes de la vía de síntesis de capsaicinoides y de rutas metabólicas relacionadas con el desarrollo que pudieran afectar su contenido en frutos de chile habanero.

3.2. METODOLOGÍA

3.2.1. Condiciones de crecimiento de las plantas

Se usaron plántulas de chile habanero (*C. chinense* Jacq.) variedad Jaguar, de aproximadamente 45 días de crecimiento, que se trasplantaron a bolsas de polietileno con 12 kg de suelo negro o suelo rojo, colocando 15 plántulas por tipo de suelo. Las plántulas fueron cultivadas en condiciones de invernadero en el Centro de Investigación Científica de Yucatán durante los meses de enero a junio de 2019, durante un ciclo de cultivo de 150 DPT, distribuyéndose por bloques de tres macetas de un mismo tipo de suelo y alternando suelo negro y suelo rojo, como se muestra en la Figura 3.1.

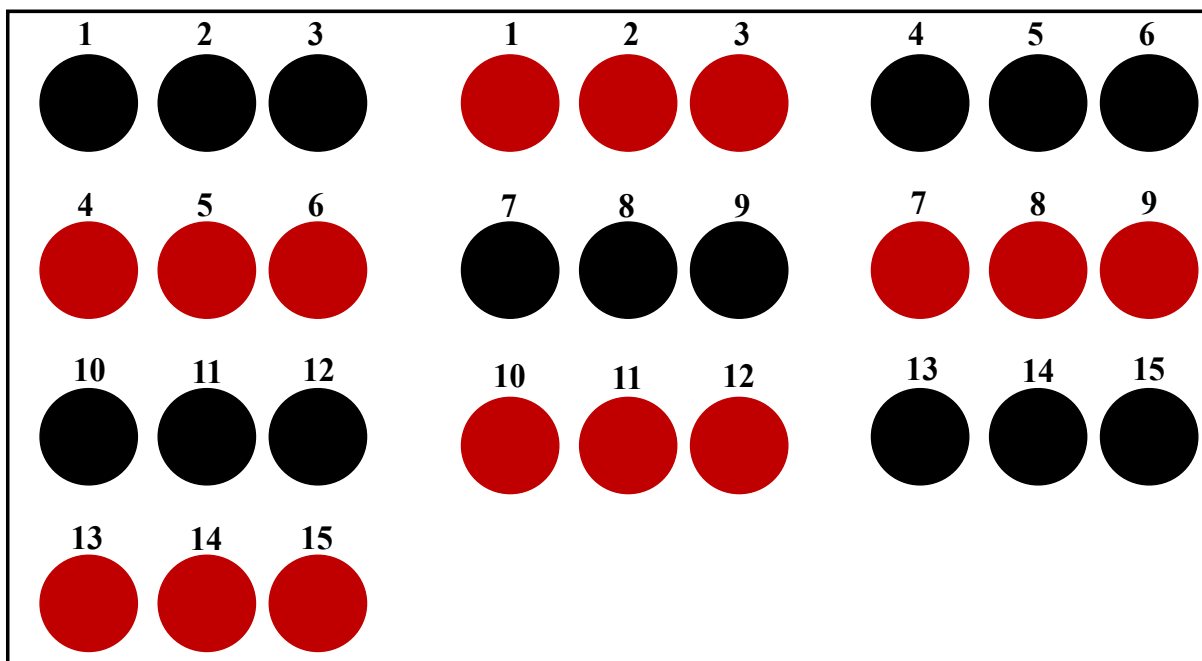


Figura 3.1. Disposición de las plantas de chile habanero en el invernadero durante el experimento. Los círculos negros representan las plantas crecidas en suelo negro y los círculos rojos representan las plantas crecidas en suelo rojo

Las plántulas fueron regadas periódicamente con la misma cantidad de agua (de 0.5 a 1.5 L), según las necesidades de las plantas. La fertilización se realizó usando Triple-18 (0.7713 g por planta) con dos aplicaciones semanales (cada tres o cuatro días). También se aplicó un día a la semana el fertilizante Bayfolan de manera foliar (3 a 6 mL por las 30 plantas). Durante la etapa de floración y del etiquetado de flores se usó el fertilizante Biozyme (750 μ L por las 30 plantas), aplicándose de manera foliar cada quince días.

3.2.2. Etiquetado de flores, colecta y caracterización de frutos

Las flores se empezaron a etiquetar a partir de que el 50% de las plantas tuvieran flores en antesis. El etiquetado se realizó diariamente en flores en antesis durante un período de 60 días. Se anotó el número de flores y de frutos colectados diariamente, a partir de los cuales se calculó el aborto floral.

Los frutos para la transcriptómica se cosecharon en etapa de 25 DPA, mientras que los frutos para la determinación de capsaicinoides se cosecharon en etapas de 25, 45 y 60 DPA. Se realizaron por lo menos tres colectas por cada punto. Los frutos se congelaron en nitrógeno

líquido y se almacenaron a -80°C . Para la caracterización de los frutos (peso y largo de frutos) se recolectaron diariamente aquellos que tuvieran 45 DPA. Los frutos frescos se pesaron en balanza analítica, mientras que el largo de los frutos se midió con un vernier. Posteriormente se almacenaron a -20°C .

3.2.3. Determinación de capsaicinoides

Se tomaron dos grupos de 5 frutos de 25, 45 y 60 DPA que se trituraron en licuadora en nitrógeno líquido, para después almacenarse a -80°C . Cada grupo se procesó por separado, teniendo dos réplicas biológicas por cada suelo y cada tiempo. Se tomaron alícuotas y se liofilizaron a -40°C en vacío durante dos días. Para la extracción de capsaicinoides, 0.15 g de muestra liofilizada se mezclaron con 20 mL de etanol absoluto anhidro y se incubó durante dos horas en matraces Erlenmeyer de 50 mL en agitador orbital a aproximadamente 75-100 rpm a temperatura ambiente. Se decantaron los restos sólidos y el sobrenadante se trasvasó a tubos Falcon. Se tomaron 1.4 mL de este sobrenadante que se centrifugaron a 10000 rpm por 10 min a 4°C . De éstos, se tomaron 500 μL de sobrenadante que se secaron en una centrífuga con vacío (VacufugePlus, Eppendorf). Una vez seca, las muestras se disolvieron en 100 μL de etanol absoluto grado HPLC.

Para la determinación de capsaicinoides se utilizó la técnica de HP-TLC con la metodología siguiente (Figura Suplementaria 3.1). Se aplicaron 5 μL de muestra sobre una placa de sílica gel 60F 254 de 20 de ancho x 10 cm de largo, usando un equipo semiautomático Linomat-5 (CAMAG), en una banda de 5.0 mm a una distancia de 10.0 mm del borde inferior de la placa. La placa se corrió con 10 mL de una fase móvil de ciclohexano:cloroformo:ácido acético (70:20:10) hasta una distancia de 70.0 mm del borde inferior de la placa en una cámara CAMAG sin saturación. Los capsaicinoides se cuantificaron por densitometría a 282 nm en un equipo CAMAG TLC Scanner 4, para ello se usaron estándares de capsaicina para identificar la banda de capsaicina de las muestras y para construir una curva de calibración en un rango de 0.25 a 3.0 μg , a partir de una solución de 1.0 $\mu\text{g}/\mu\text{L}$ de capsaicina, y que se aplicaron sobre las placas de sílica gel con la misma metodología que con las muestras. Para cada réplica de frutos de 25 y 45 DPA, la aplicación y cuantificación de capsaicinoides se realizó por quintuplicado, mientras que para los de 60 DPA, se realizó con 6 repeticiones por réplica.

3.2.4. Análisis estadístico

El número de flores, número de frutos, aborto floral, peso total de frutos y peso promedio de frutos se analizaron por ANOVA de una vía usando la prueba de Tukey con $n=8$ plantas para el número de flores y frutos totales, $n=13$ plantas para el aborto floral, y $n=7$ plantas para las comparaciones de frutos de 45 DPA. La determinación de capsaicinoides se realizó por ANOVA de dos vías, con los factores: tipo de suelo (con los niveles suelo negro y suelo rojo) y estado de desarrollo (con los niveles 25, 45 y 60 DPA), con $n=10$ frutos para las comparaciones en 25 y 45 DPA, y $n=12$ frutos para las comparaciones de 60 DPA. El nivel estadístico de significancia fue ajustado a $P=0.05$. Se usó el software SigmaPlot v.12.0 para realizar los ANOVA y para construir las gráficas de barras para comparar los diferentes parámetros.

3.2.5. Extracción de RNA

Para la extracción se utilizaron dos grupos de 5 frutos de 25 DPA para suelo negro y suelo rojo con el objetivo de tener dos réplicas biológicas por tipo de suelo, y que se procesaron en licuadora por separado de manera similar a como se realizó para los frutos usados en la determinación de capsaicinoides.

Con los frutos pulverizados se realizó la extracción de RNA usando el kit RNeasy Plant Minikit (Qiagen) con la metodología siguiente. Se tomaron 0.1 g de muestra que se molieron en mortero en presencia de nitrógeno líquido. Se adicionaron 450 μL de buffer RLT con 10 μL de β -mercaptoetanol por 1 mL de buffer RLT, y se continuó pulverizando la muestra hasta homogeneizar la muestra con el líquido.

El lisado se transfirió a las columnas spin QIAshredder, donde se centrifugó 2 min a 13000 rpm a 4 °C. El sobrenadante del filtrado se transfirió a tubos Eppendorf donde se adicionaron 0.5 volúmenes de etanol absoluto. Posteriormente, se transfirió la muestra a tubos RNeasy Spin Column, donde se centrifugó a 10000 rpm 1 min a 4° C. Se descartó el filtrado y en la columna se adicionaron 700 μL de buffer RW1, se centrifugó a 10000 rpm, 1 min a 4° C, se descartó el filtrado. Se adicionaron 500 μL de buffer RPE, y se realizó otra centrifugación a 10000 rpm, 4 ° C durante 1 min. Este paso se realizó dos veces. Finalmente, a las columnas se les adicionó 40 μL de agua libre de nucleasas y se centrifugó a 11000 rpm durante 3 min para eluir el RNA en tubos Eppendorf.

Se tomaron 6 µg de RNA total de las extracciones y se trataron con Turbo DNase (2 U/µL, Invitrogen) a 37° C por 20 min siguiendo el protocolo establecido por el fabricante. Los resultados de las extracciones se observaron por electroforesis en gel de agarosa 1.5% a 90 V. Las cuantificaciones de RNA se realizaron en nanodrop a 260 nm, y se midieron los coeficientes 260/280 y 260/230.

3.2.6. Secuenciación del transcriptoma de frutos de chile habanero

Los extractos de RNA total libres de DNA se trataron con el kit RNAsable/RNAsable LD (Biomatrica) siguiendo el protocolo recomendado por el fabricante. Las muestras se homogeneizaron en los tubos RNAsable, homogeneizándose con el contenido incluido en éstos y se secaron dentro de los tubos en una centrifuga con vacío (VacufugePlus, Eppendorf) por 60-90 min. Una vez secas, las muestras se guardaron dentro de bolsas con sílica gel del kit RNAsable. Estas muestras se enviaron a MacroGen para realizar la secuenciación del transcriptoma.

La secuenciación se realizó en la plataforma Illumina por TruSeq, esta se realizó para cada réplica biológica de suelo negro y suelo rojo de los frutos de 25 DPA, teniendo en total cuatro librerías (SN1, SN2, SR1 y SR2) por *paired ends*. Brevemente, para la preparación de las librerías, se purificó el mRNA de las muestras de RNA total mediante perlas magnéticas unidas a oligos de poliT, se fragmentó el mRNA y se sintetizó el cDNA de estos fragmentos por transcripción inversa usando hexámeros aleatorios. Una vez realizada la síntesis de la doble cadena de cDNA, a estos fragmentos se les ligó adaptadores en ambos extremos, y se amplificaron por PCR. Se seleccionaron aquellos fragmentos con tamaños de 200 y 400 pb para formar las librerías, que se secuenciaron para obtener *reads* pareados (*paired-ends*) de 101 bases.

3.2.7. Filtrado de *reads* y ensamblado

El *trimming*, ensamblado, predicción de marcos de lectura abiertos (ORFs), anotación funcional de la biblioteca de *contigs* y estimación de la abundancia de los unigenes se realizó con los servicios de MacroGen. El filtrado de *reads* se realizó mediante Trimmomatic v.0.38 (Bolger *et al.*, 2014), eliminado adaptadores y bases de calidad baja (<Q3) de los extremos de estos. Además, se usó una ventana (*sliding window*) de 4 bases para eliminar bases por debajo de una calidad promedio de Q15. Los *reads* resultantes con una longitud inferior a 36 bases fueron eliminados. Para observar y evaluar los resultados de calidad de los transcriptomas se usó el

software FastQC v.0.11.7 (Andrews, 2010). Tras el filtrado, se realizó el ensamblado *de novo* utilizando el programa Trinity, versión trinityrnaseq_r20140717, con los parámetros por default (Haas *et al.*, 2013). Para realinear los *reads* en el transcriptoma ensamblado, se utilizó el software Bowtie 1.1.2 (Langmead *et al.*, 2009). Los *contigs* fueron filtrados y agrupados en transcritos no redundantes, que se definieron como “unigenes”, usando el software CD-HIT v.4.6. (Fu *et al.*, 2012).

3.2.8. Predicción de ORFs, anotación funcional y estimación de la abundancia de los unigenes

La predicción de los ORFs de al menos 100 aminoácidos de longitud en los unigenes se realizó utilizando el programa TransDecoder v.3.0.1 (<https://github.com/TransDecoder/TransDecoder/wiki>) para identificar regiones codificantes candidatas en las secuencias de los transcritos. Para la estimación de la abundancia de los genes, primero los *reads* fueron alineados a la biblioteca de unigenes con el software Bowtie 1.1.2, y luego la abundancia de los transcritos fue estimada como conteo de *reads*, en FPKM (fragmentos por kilobase de transcrito por millones de *reads* mapeados) y en TPM (transcritos por millón) usando el programa RSEM v.1.3.1. (Li y Dewey, 2011). La anotación funcional de los unigenes se realizó contra las bases de datos GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes) y la base de datos de nucleótidos (nt) y de proteínas no redundante (nr) del NCBI, usando blastn (BLAST v.2.9.0) para la base nt y blastx con el software DIAMOND v.0.9.21 (Buchfink *et al.*, 2021) para las otras tres bases de datos, usando un valor e límite de máximo 1.0 e-5. Para las anotaciones GO, éstas se clasificaron respecto a cada una de las tres categorías funcionales principales: proceso biológico, componente celular y función molecular. En el caso de la anotación contra KEGG, la búsqueda se realizó con el método BBH (Bi-directional best hit).

3.2.9. Análisis de expresión diferencial y de rutas metabólicas enriquecidas de los unigenes

El análisis expresión diferencial de genes (DGE) de la biblioteca de unigenes se realizó utilizando el programa DESeq2 (Love *et al.*, 2014) comparando la expresión de los genes de suelo rojo contra suelo negro. Se estableció un límite de \log_2 Fold Change (LFC) ≥ 1.5 para los genes más expresados en suelo rojo y un valor LFC ≤ -1.5 para los genes más expresados en suelo negro. En ambos casos se estableció un valor p < 0.00835 como límite de confianza.

Para la representación de la expresión de los genes se construyó una gráfica de volcán de LFC contra los valores p utilizando el paquete ggplot2 (Wickham, 2016). Posteriormente se realizó un análisis de enriquecimiento de sets de genes (GSEA) usando las anotaciones KEGG de *C. annuum*. Primero se sumaron los *reads* de todos los *contigs* que estuvieron anotados con el mismo término KEGG y luego se realizó el DGE mediante DESeq de manera similar a como se explicó anteriormente. Tras esto, se realizó el GSEA mediante ClusterProfiler v.4.4.4 (Wu *et al.*, 2021) con los paquetes ggplot2, enrichplot (Yu, 2022) y pathview (Luo y Brouwer, 2013). El GSEA se corrió con 10000 permutaciones, un tamaño de grupo de gen (gene set) de mínimo 3 y máximo 800, y un límite de valor p de 0.5. Para el uso de los programas y paquetes anteriores se utilizó R v.4.2.0.

3.2.10. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides

Para la selección de genes de las rutas metabólicas del análisis GSEA y genes representativos de la ruta de síntesis de capsaicinoides se realizó una búsqueda de transcritos usando secuencias CDS (secuencias codificantes) de genes de referencia contra la biblioteca de unigenes mediante un *blastn* local mediante el programa Bioedit, con un valor *e* de 0. Se utilizaron como anzuelos genes regulados de las rutas metabólicas enriquecidas con homólogos de *C. annuum* encontrados en la base de datos KEGG tras el análisis GSEA. Como genes anzuelos en el metabolismo de capsaicinoides se utilizaron homólogos reportados como regulados para la ruta de biosíntesis de capsaicinoides. Para los genes *AT3-D1* y *AT3-D2* (Qin *et al.*, 2014) se usaron las secuencias CDS anotadas en el cromosoma 2 del genoma de *C. chinense* (MCIT02000002.1) de las proteínas PHU26684.1 y PHU26685.1, respectivamente. Para *pAMT* se utilizó la secuencia CDS AF085149.1 (Tanaka *et al.*, 2018). Para *KAS* y *FAT*, se utilizaron los CDS AF085148.1 y AF318288.1, respectivamente (Sun *et al.*, 2022). Para *KR* se usó la secuencia CDS EU616561.1 (Koeda *et al.*, 2019). Para *MYB31* se usaron los CDS MK074771.1, MK074772.1 y MF062086.1 (Sun *et al.*, 2022; Han *et al.*, 2019; Zhu *et al.*, 2019). Se construyó un mapa de calor para analizar la expresión de estos genes en unidades TPM utilizando los paquetes gplots (Warnes *et al.*, 2022) y RColorBrewer (Neuwirth, 2022) en R v.4.2.0.

3.3. RESULTADOS

3.3.1. Producción de flores y frutos y caracterización de frutos en plantas de chile habanero crecidas en dos tipos de suelo

Se cultivaron plántulas de chile habanero de la variedad Jaguar en suelo negro (SN) y suelo rojo (SR) durante un ciclo de 150 DPT. Durante el primer ciclo de floración (60 días) se midió el número de flores, así como la producción de frutos correspondiente comparando ambos tipos de suelo (Figura 3.2). No se encontraron diferencias significativas en el número de flores totales. Los números fueron de 305 y 284 flores para SN y SR, respectivamente, pero el número de frutos totales a los 150 días DPT se mantuvo entre 40 y 35 en SN y SR, respectivamente, sin encontrarse diferencias significativas entre ellos. Si tomamos en cuenta el porcentaje de aborto floral, que fue del 85% para los dos tipos de suelos, este número de frutos es bajo. Este comportamiento es bastante común en esta especie cultivada en las condiciones de Yucatán donde las temperaturas están cercanas a los 40° C en verano, aun siendo esta especie bastante tolerante al déficit hídrico (Nieto-Garibay *et al.*, 2022). El número de frutos de 45 DPA por suelo no superó los 40. Si tomamos en cuenta el número de frutos totales, no se encontraron diferencias entre estos parámetros en los diferentes tipos de suelo (Figura 3.2A).

Un parámetro importante para el estudio del desarrollo de los frutos en esta especie es determinar los rangos de tamaño de los frutos (Figura 3.2B). Estos rangos fueron medidos en frutos de 45 DPA, fecha en la cual ya comienza la fase de maduración. La mayor cantidad de frutos de esta etapa tuvo un tamaño entre 3 y 3.99 cm, con un número de frutos entre 19 y 16 en SN y SR, respectivamente, sin mostrar diferencias significativas. Los frutos más grandes (con un tamaño de 4 cm de largo o superior) fueron menos en número, entre 3 y 7 en SN y SR, respectivamente; sin embargo, se encontró que el tamaño era significativamente mayor en SR. También se encontró diferencia significativa en los frutos de 2 a 2.99 cm en SN y SR, respectivamente, teniendo mayor número de frutos las plantas crecidas en SN. El tamaño de frutos más pequeños (menores a 2 cm de largo) no fue significativamente diferente en los diferentes tipos de suelo.

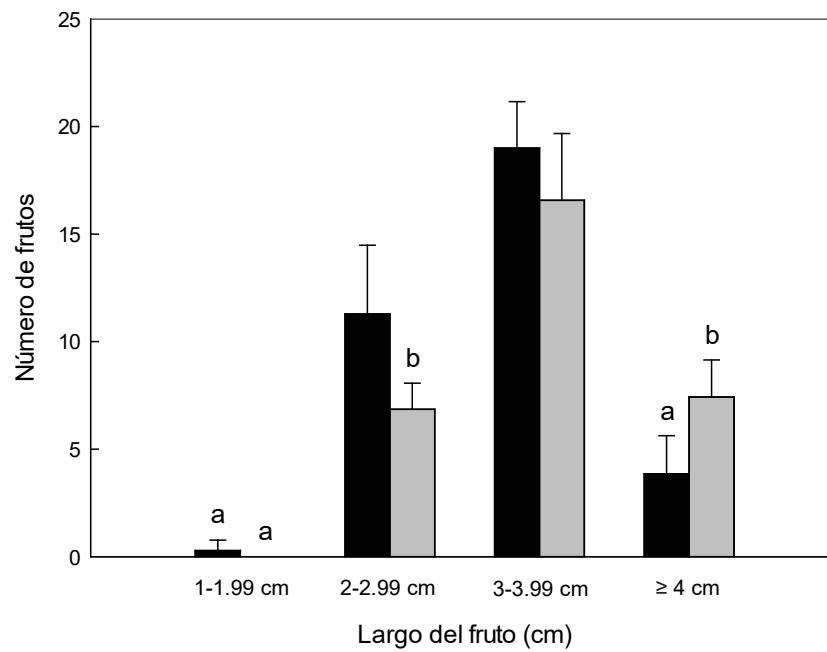
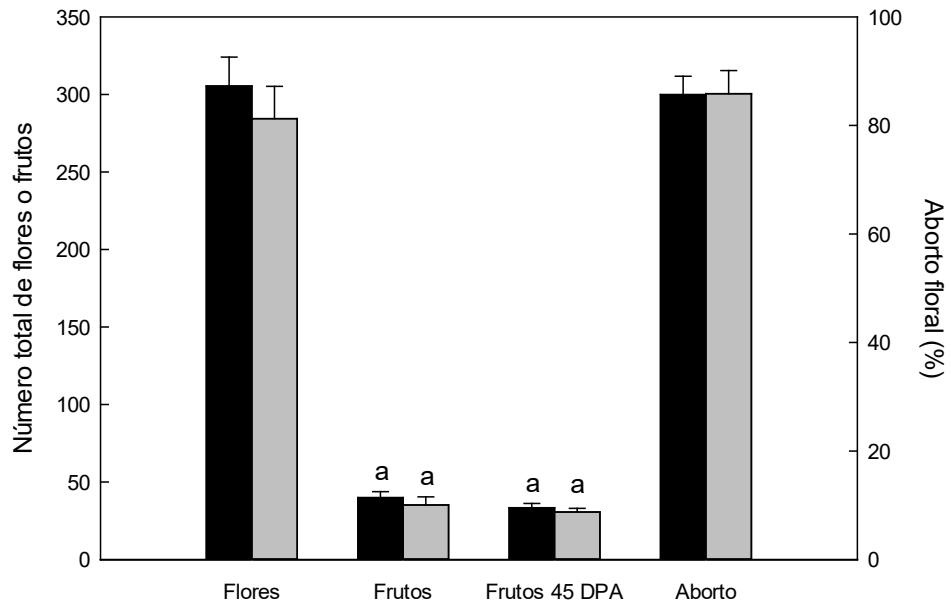
Si tomamos en cuenta el peso total promedio por planta de los frutos de 45 DPA (Figura 3.2C), se observa que los frutos cuyas medidas estuvieron en los rangos de 2-2.99 cm y de 3-3.99 cm en SN y SR fueron los más pesados, teniendo para el primer rango 36 y 22 g, y en el segundo 94 y 73 g en SN y SR, respectivamente, y encontrándose diferencias significativas entre ellos,

CAPÍTULO III

coincidiendo en que en ese rango está el mayor número de frutos en cada tratamiento (Figura 3.3B), lo cual es un comportamiento esperado. Los frutos con un tamaño de 4 cm o superior mostraron diferencias significativas en cuanto al peso total, siendo los frutos de SR los que tuvieron mayor peso total promedio por planta, de 51 g, mientras que los de plantas cultivadas en SN pesaron en promedio 26 g. Los frutos más pequeños, menores a 2 cm de largo, tuvieron un peso total promedio similar en ambos tipos de suelo.

Finalmente, se midió el peso promedio por fruto de 45 DPA en los diferentes tipos de suelo (Figura 3.2D). De manera esperada la tendencia fue que a medida que los frutos eran más grandes su peso se incrementara, no se encontraron diferencias significativas en ninguno de los pesos en relación al tamaño en los diferentes tipos de suelos estudiados, a excepción de los frutos de 3-3.99 cm, que tuvieron un mayor peso promedio en SN (5.30 g) en comparación a SR (5.05 g). Esto quiere decir que, en general, en ambos suelos el peso promedio fue similar para los diferentes tamaños de fruto, y permite, hasta el momento, utilizar cualquiera de los suelos para la cosecha de frutos de 45 DPA. Los pesos oscilaron entre 3 y 7 g en ambos suelos en los frutos de 2 cm o superior de largo.

Podemos concluir que tomando en cuenta los parámetros estudiados que en ambos suelos el porcentaje de aborto, el número de flores, el número de frutos totales y el peso promedio de los frutos no muestran diferencias significativas, lo que nos indica que cualquier tipo de suelo estudiado es ideal para el cultivo de esta especie, aun cuando por tamaños sí se observa una diferencia en el número total de frutos producidos, favoreciendo un mayor número de frutos más grandes en SR. Sin embargo, esta diferencia es despreciable si tomamos en cuenta que esto es parte del desarrollo de los frutos y que lo importante es la producción total.



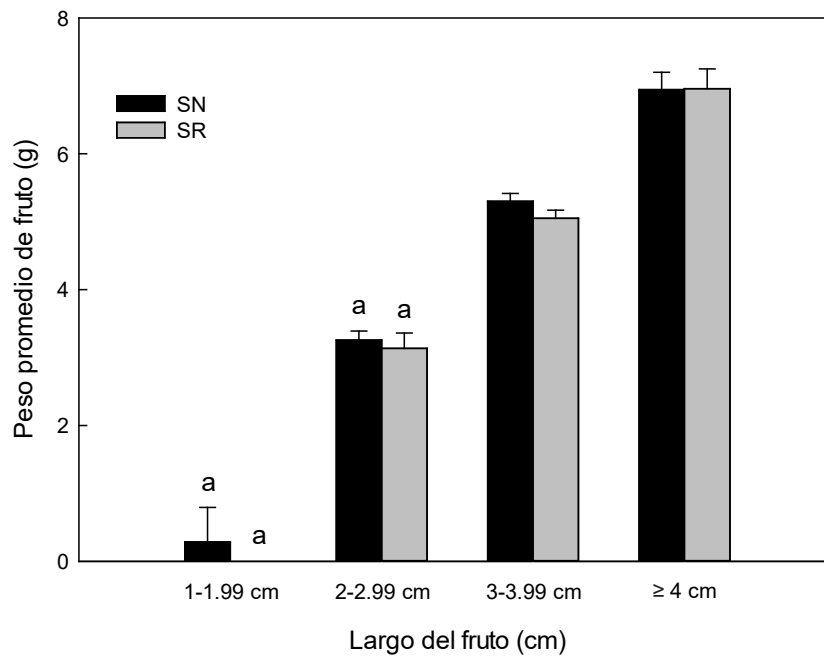
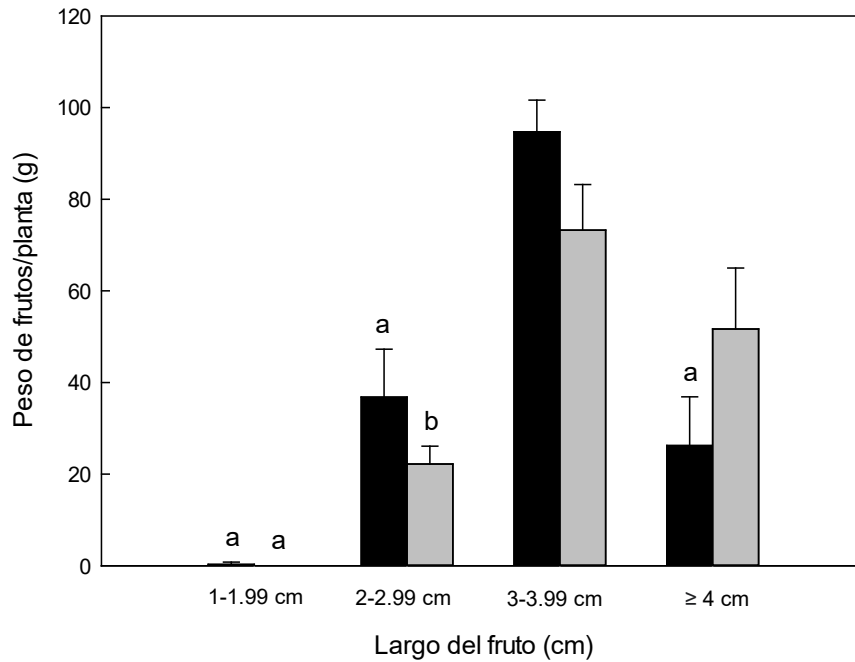


Figura 3.2. Producción de flores y frutos en plantas de chile habanero crecidas en suelo negro (SN) o suelo rojo (SR) tras 150 DPT. A) Producción total de flores, frutos, frutos de 45 DPA y aborto floral durante el ciclo de cultivo de 150 DPT, B) número total de frutos de 45 DPA divididos en rangos de tamaño durante los 150 DPT comparando plantas en ambos tipos de suelo, C) peso total de los frutos de 45 DPA, y D) peso promedio de los frutos de 45 DPA. Letras diferentes indican diferencia en el parámetro medido de acuerdo a la prueba de Tukey ($p < 0.05$), con $n=8$

para el número de flores y frutos totales, n=13 para el aborto floral, y n=7 para las comparaciones de frutos de 45 DPA

3.3.2. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en dos tipos de suelo

Uno de los parámetros más importantes en la medición de la calidad del chile habanero es la determinación del contenido de capsaicinoides. Estos compuestos son los que les da un valor agregado a los frutos de esta especie, especialmente la capsaicina que junto con la dihidrocapsaicina representan el 90 % de los capsaicinoides y les dan el picor a los frutos, lo cual es altamente apreciado por los consumidores y es un parámetro fundamental en la calidad de esta especie (Ruiz-Lau *et al.*, 2011).

En la figura 3.3 se muestran los resultados del contenido de capsaicina en frutos completos de plantas de chile habanero crecidas en suelo negro (SN) o suelo rojo (SR) en distintos estados de desarrollo (25, 45 y 60 DPA). El contenido de capsaicina se mantuvo en niveles entre 7 y 12 mg/g peso seco de fruto (PS). Se compararon los valores entre etapa de desarrollo de fruto en el mismo suelo, y además entre tipo de suelo en el mismo estado de desarrollo, para poder reportar en cuál de los suelos estudiados se encontraban los mayores contenidos, ya anteriormente se había concluido que los suelos no mostraban diferencias en los parámetros fisiológicos estudiados. Se encontró que la diferencia en capsaicina entre tipos de suelo depende del estado de desarrollo analizado. A los 25 DPA se encontraron diferencias significativas en la cantidad de capsaicina entre suelos, siendo en SN donde los frutos mostraron los mayores valores (10.72 mg/g PS), mientras que en SR fue de 8.92 mg/g PS. A los 45 DPA, no se encontraron diferencias en la cantidad de capsaicina entre suelos, aunque aumentó debido al proceso de síntesis de estos metabolitos. En cuanto a los frutos de 60 DPA, al contrario de 25 DPA, se tuvo un mayor contenido de capsaicina en SR, aunque los valores fueron menores debido al proceso de degradación de estos compuestos en estados de desarrollo avanzado de los frutos.

Cuando comparamos el contenido de capsaicina dentro del mismo suelo en las diferentes etapas de desarrollo, podemos ver que no hay diferencias significativas entre 25 y 45 DPA en SN, mientras que en SR sí se encontraron diferencias entre los tiempos señalados, con valores de 8.92 y 11.48 mg/g PS de fruto para 25 y 45 DPA, respectivamente. Esto indica que los frutos no sobrepasaron un máximo de capsaicina independientemente del tipo de suelo en el que se

cultivó la planta. Además, este máximo se alcanzó antes en SN, mientras que el SR la síntesis continua hasta que se alcanza la etapa en la que los frutos empiezan a madurar (45 DPA). Posteriormente, debido a la degradación de capsaicinoides, a los 60 DPA hubo una disminución en el contenido de estos en ambos tipos de suelo, aunque este proceso fue afectado por el tipo de suelo, donde en SN la degradación fue superior, teniendo un contenido de capsaicina por debajo del que había a los 25 DPA (7.72 contra 10.72 mg/g PS de fruto).

Podemos concluir que si colectamos frutos de 45 DPA no es importante el tipo de suelo donde se cultive esta especie si lo hacemos con la finalidad de obtener capsaicina para la industria o en su caso frutos picosos para el consumo en fresco. Tanto los suelos negros como los rojos son adecuados para el cultivo de chile habanero, solo es necesario saber cuál es la finalidad de esos frutos, si son para la industria o para el consumo fresco, las variaciones encontradas en el contenido de capsaicinoides fueron debidas a las diferentes etapas de crecimiento.

También es importante considerar que los resultados sugieren un efecto en la síntesis de los capsaicinoides por el tipo de suelo por el cual su contenido alcanzó un máximo en frutos de 25 DPA en SN mientras que en SR continuó la acumulación hasta los 45 DPA. Este efecto podría verse reflejado a nivel transcriptómico, por lo que se procedió a analizar la expresión de genes relacionados con la síntesis de capsaicinoides en los frutos de 25 DPA.

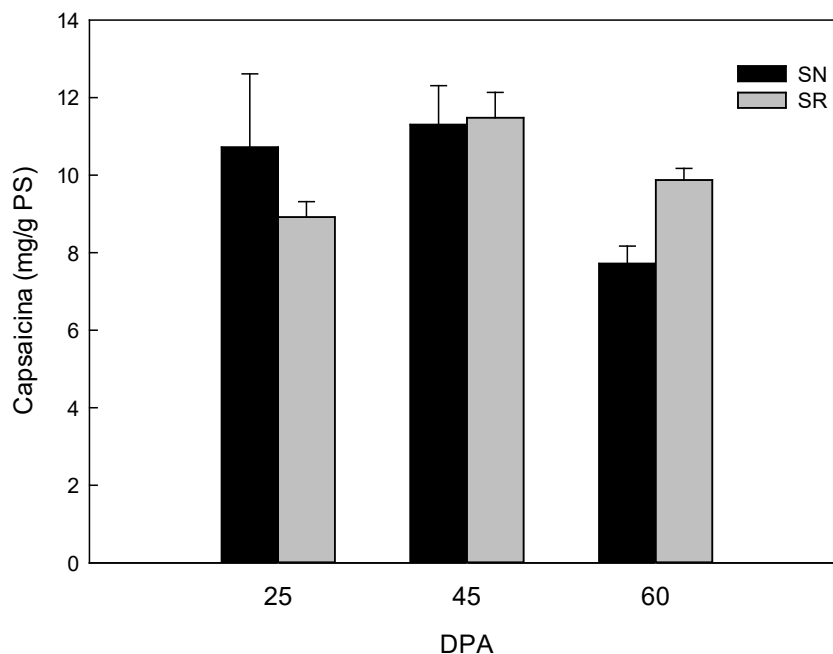


Figura 3.3. Contenido de capsaicinoides en frutos completos de plantas de chile habanero crecidas en suelo negro (SN) o suelo rojo (SR) en distintos estados de desarrollo (25, 45 y 60 DPA). Los resultados se expresan como mg de capsaicina/g de peso seco (PS) de fruto. Letras minúsculas distintas indican diferencia entre tiempos para el mismo tipo de suelo y letras mayúsculas distintas indican diferencia entre suelos para el mismo tiempo de acuerdo a la prueba de Tukey, con n=10 en frutos de 25 y 45 DPA y n=12 para frutos de 60 DPA

3.3.3. Análisis general del transcriptoma y de rutas metabólicas enriquecidas en frutos de plantas de chile habanero crecidas en dos tipos de suelo

3.3.3.1. Análisis del general del transcriptoma

Se analizó el transcriptoma de frutos de chile habanero de 25 DPA. Este tiempo fue elegido debido a que el contenido de capsaicinoides en los frutos de chile habanero fue diferente entre los tipos de suelo, donde en SN la acumulación de capsaicinoides alcanza su máximo, mientras que en SR está síntesis continúa (Figura 3.2). Además, en estas etapas intermedias de desarrollo se conoce que la síntesis de capsaicinoides está activa en frutos y la expresión de los genes de la vía de síntesis se encuentra al alza (Zhang *et al.*, 2016). Se generaron cuatro bibliotecas de cDNA en *paired-ends* dos de SN y dos de SR, secuenciadas a partir de dos grupos de 5 frutos por cada tipo de suelo. Tras el filtrado se tuvieron entre 42.6 a 55.9 millones de *reads* totales. Entre 95-96% de los *reads* tuvieron una calidad por base igual o por encima de Q30, y la calidad promedio de *reads* de alrededor de Q36 (Figuras Suplementarias 3.2 y 3.3 y Tabla 3.1), indicando una alta calidad de las secuencias en las cuatro bibliotecas. La longitud de los *reads* estuvo entre 36-101 bases, donde la mayor parte de las secuencias tuvieron 101 bases (Figura Suplementaria 3.4). El tamaño de las bibliotecas estuvo entre 4289 y 5625 Mb. La composición GC fue similar entre muestras, lo cual era esperado ya que pertenecen a frutos del mismo tiempo de desarrollo y la misma variedad (Tabla 3.1)

Tabla 3.1. Estadísticas de los *reads* filtrados del transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro (SN) y suelo rojo (SR), realizando dos réplicas (1 y 2) por tipo de suelo

Muestra	Bases totales	Reads totales	Longitud (bases)	GC (%)	*Q20 (%)	**Q30 (%)
SN1	4289416937	42654572	36-101	43.95	98.99	96.16
SN2	5625995957	55920028	36-101	44.06	98.98	96.12
SR1	5049889977	50383680	36-101	43.16	98.74	95.63
SR2	4833956215	48192058	36-101	43.40	98.81	95.78

CAPÍTULO III

*Q20: porcentaje de bases con una calidad de valor phred igual o superior a 20. Esto es la probabilidad de encontrar 1 base incorrecta cada 100 bases, o una precisión de 99%

**Q30: porcentaje de bases con una calidad de valor phred igual o superior a 30. Esto es la probabilidad de encontrar 1 base incorrecta cada 1000 bases, o una precisión de 99.9%

Tras el ensamblado *de novo* y el filtrado de secuencias redundantes, se obtuvo un transcriptoma de 99159 unigenes con un promedio de longitud de 689.55 pb, una composición GC de 38.95% y 68.37 Mb. La longitud mínima de *contig* fue 201 pb y la máxima fue 9421 pb, y el N50 fue de 1143 pb (Tabla 3.2.A). Valores similares de ensamblado *de novo* de transcriptomas han sido obtenidos en frutos inmaduros (20-25 DPA), en estado intermedio de maduración (30-40) y completamente maduros (50-60 DPA) en otras variedades picantes de *C. chinense* y en *C. frutescens* (Chhapekar *et al.*, 2020). Además, es importante conocer cuántos *reads* son realineados al transcriptoma ensamblado. El porcentaje de *reads* mapeados al transcriptoma de unigenes estuvo entre 78 a 82% (Tabla 3.2B), lo cual indica que una gran cantidad de la información de los *reads* se encuentra en el ensamble, que permite que se pueda usar para encontrar genes expresados y analizar los niveles de expresión en las cuatro bibliotecas de los frutos (Chhapekar *et al.*, 2020; Zhang *et al.*, 2016).

Tabla 3.2. Resumen del ensamble del transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro (SN) y suelo rojo (SR). Los resultados corresponden a los *contigs* no redundantes obtenidos tras filtrado y agrupamiento con CD-HIT. Se muestran las estadísticas de los parámetros de ensamblado (A) y del realineamiento de los *reads* filtrados de cada biblioteca en el transcriptoma ensamblado (B)

A	
Parámetro	Contig "unigene"
"Unigenes" totales	99159
%GC	38.95
*N90 (pb)	274
*N50 (pb)	1143
*N10 (pb)	3195
Longitud máxima de <i>contig</i> (pb)	9421
Longitud mínima de <i>contig</i> (pb)	201
Mediana de la longitud de <i>contig</i> (pb)	379.0
Promedio de la longitud de <i>contig</i> (pb)	689.55
Bases totales ensambladas	68375141

B		
Muestra	Reads mapeados (%)	Reads no mapeados (%)
SN1	34696414 (81.34)	7958158 (18.66)
SN2	46129206 (82.49)	9790822 (17.51)
SR1	39660592 (78.72)	10723088 (21.28)
SR2	38558338 (80.01)	9633720 (19.99)

*Nx: este valor indica que un x% de la información (nucleótidos) se encuentran en *contigs* de al menos cierta longitud, cuando los *contigs* ensamblados se ordenan de mayor a menor tamaño. El N50 indica que al menos el 50% de la información del transcriptoma está representada en *contigs* de al menos 1143 pb

3.3.3.2. Análisis de genes diferencialmente expresados y de rutas metabólicas enriquecidas entre los frutos de chile habanero en suelo negro y suelo rojo

Se realizó la anotación de las cuatro bibliotecas contra la base de datos GO, agrupando las subcategorías funcionales más abundantes de las tres categorías principales: proceso biológico (BP), componente celular (CC) y función molecular (MF) (Figura 3.4). Un total de 30993 unigenes del transcriptoma ensamblado fueron anotados en la base de datos GO. Los cuatro transcriptomas tuvieron similar proporción en las subcategorías dominantes.

Dentro BP las categorías dominantes fueron: “proceso celular” (15910 unigenes) y “proceso metabólico” (13790 unigenes), seguido de “regulación biológica” (7048 unigenes) y “respuesta a estímulos” (5345 unigenes). En CC las categorías mayoritarias fueron: “parte celular” (24747 unigenes) y “organelo” (15865 unigenes), seguido de “membrana” (6602 unigenes) y “parte de organelo” (6304 unigenes) en similar proporción. En cuanto a MF, “actividad catalítica” y “unión” fueron las categorías con mayor proporción de unigenes (14398 y 12176, respectivamente), seguidos de “actividad de regulación de la transcripción” (1892 unigenes) y “actividad transportadora” (1878 unigenes) (Figura 3.4).

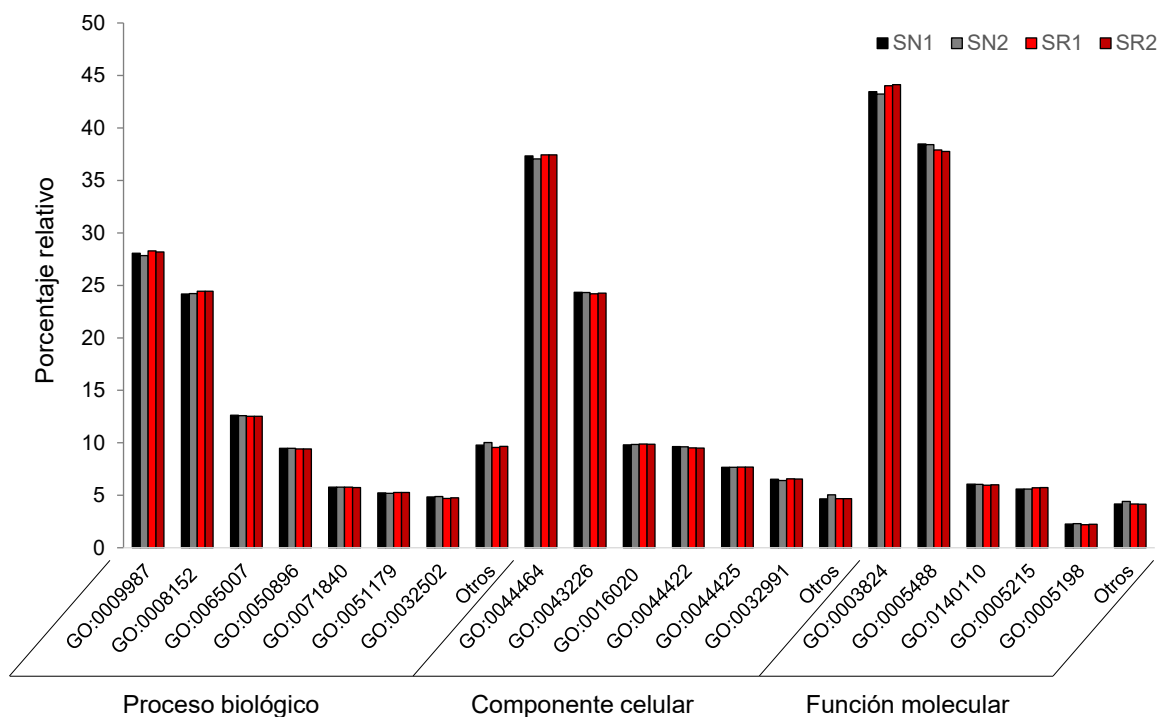


Figura 3.4. Análisis de las categorías funcionales principales GO de los *contigs* de frutos de plantas de chile habanero crecidas en SN o SR. El porcentaje relativo indica el número de *contigs* dentro de la misma categoría funcional principal (proceso biológico, componente celular o función molecular) para cada réplica (SN1, SN2, SR1 o SR2). En la categoría “Otros” se agruparon subcategorías con menos del 3% (en proceso biológico y componente celular) o menos del 1% (en función molecular) de *contigs*. GO:0009987: proceso celular, GO:0008152: Proceso metabólico, GO:0065007: regulación biológica, GO:0050896: respuesta a estímulos, GO:0071840: organización de componentes celulares o biogénesis, GO:0051179: localización, GO:0032502: proceso de desarrollo, GO:0044464: parte celular, GO:0043226: organelo, GO:0016020: membrana, GO:0044422: parte de organelos, GO:0044425: parte de membranas, GO:0032991: complejo con proteínas, GO:0003824: actividad catalítica, GO:0005488: unión, GO:0140110: regulación de la transcripción, GO:0005215: actividad de transportador, GO:0005198: actividad de molécula estructural

Para comparar cuantos genes se encontraban mayormente expresados entre tipos de suelo, se realizó un análisis DGE. De los 99159 unigenes iniciales, 6292 tuvieron cero cuentas en las cuatro bibliotecas (SN1, SN2, SR1 y SR2), por lo que se descartaron en el análisis. Los 92867 *contigs* restantes se analizaron con ayuda de una gráfica de volcán (Figura 3.5A), que permite comparar la expresión en función del valor de LFC, que expresa cuantas veces está más expresado un gen en un tipo de suelo, y el valor p, que expresa la significancia estadística de la comparación (McDermaid *et al.*, 2019).

Debido a que la comparación se realizó como SR contra SN, los valores positivos de LFC representaron genes más expresados en frutos de SR, y los valores negativos representaron aquellos genes con mayor expresión en frutos de SN. Un total de 533 unigenes tuvieron una expresión diferencial entre ambos suelos, 83 tuvieron mayor expresión en frutos de suelo negro ($LFC \leq -1.5$ y $pvalue < 0.00835$) y 450 tuvieron mayor expresión en frutos de suelo rojo ($LFC \geq 1.5$, $pvalue < 0.00835$) (Figura 3.5A). Entre los genes con mayor expresión en SR se encontraron las proteínas F-box, extensinas, proteoglicano rico en prolina, poliubiquitina, dominio de proteína BURP, proteína cinasa similar a receptor de repetición rico en leucina, proteína mago nashi, calmodulina-lisina N-metiltransferasa y 3-cetoacil-CoA sintasa (Tabla Suplementaria 3.1A).

Entre los genes más expresados en frutos de SN se encontraron ortólogos de RuBisCO (Ribulosa-1,5-bifosfato carboxilasa/oxigenasa), proteínas de unión a clorofila, BREVIS RADIX, proteínas de mitocondria, similar a albúmina y proteína de dedo de anillo-H2 (Tabla Suplementaria 3.1B).

Posteriormente se analizaron los genes en sets en función de sus rutas metabólicas para encontrar vías que estuvieran influenciadas por el tipo de suelo. Estos análisis de genes en grupos tienen las ventajas de facilitar el análisis del comportamiento de la expresión de los genes de grandes bases de datos en categorías o sets que permiten analizar tendencias de los genes dentro de esas categorías y que puedan tener un significado biológico como respuesta a factores de interés (Martínez-López *et al.*, 2014).

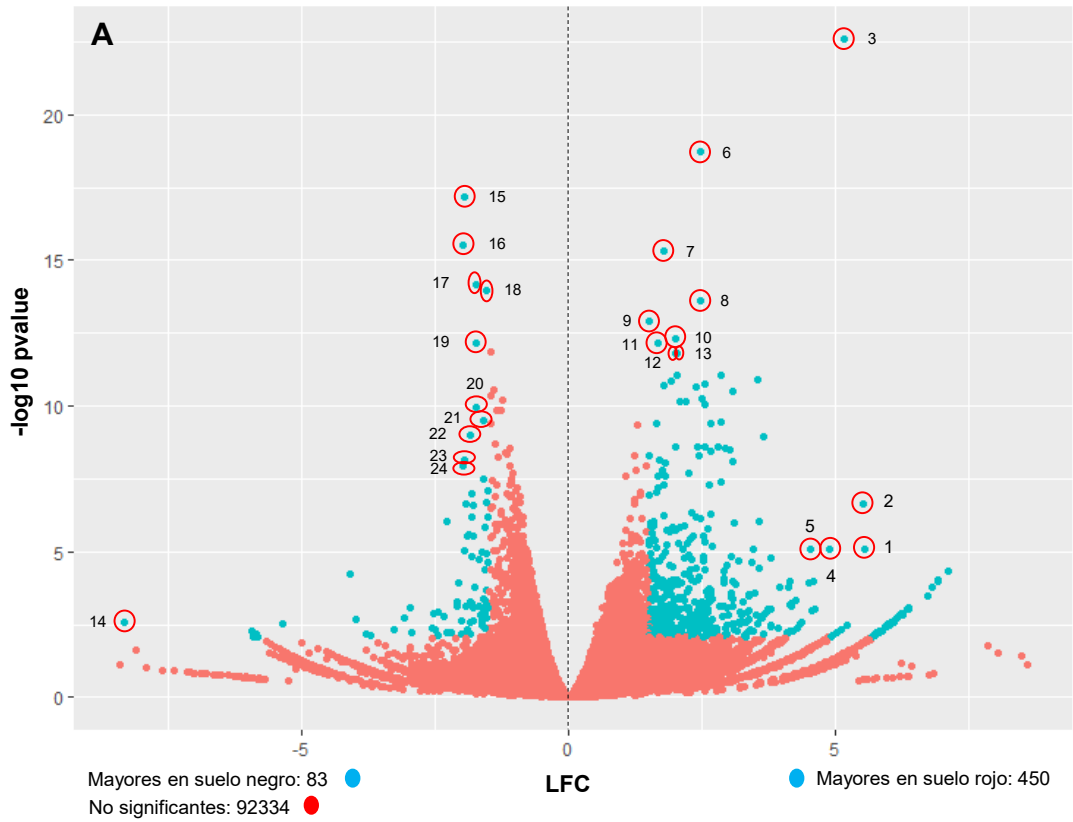
Para este fin se realizó un análisis de enriquecimiento con la anotación de los unigenes del transcriptoma de los frutos de chile habanero por términos KEGG. De los 99159 unigenes, 42936 fueron anotados en con la base de datos de KEGG. Se realizó un GSEA con los DGE, en donde previamente se sumaron los *reads* de todos aquellos *contigs* anotados con el mismo término KEGG, lo que permite encontrar rutas metabólicas sobrerrepresentadas por los valores de LFC en un tipo de suelo contra el otro (Wu *et al.*, 2021; Nguyen *et al.*, 2019). En la Figura 3.5B se muestran 11 rutas metabólicas enriquecidas, con una significancia estadística alta (*pvalue*). Estas rutas fueron de proteínas antena de fotosíntesis, la ruta de las pentosas fosfato, ritmo circadiano, biosíntesis de antocianinas, de terpenoides, de carotenoides, de arginina, metabolismo de alanina, aspartato y glutamato, del ácido 2-oxocarboxílico, de la porfirina, y del proteasoma.

CAPÍTULO III

La ruta metabólica con mayor número de genes fue la ruta de las pentosas fosfato, mientras que la biosíntesis de antocianinas tuvo el menor número de genes. La vía con mayor proporción de genes en un suelo con respecto al total de genes anotados en esa ruta (GeneRatio) fue la de las proteínas antena de fotosíntesis, mientras que la que menor proporción tuvo fue la del metabolismo de la porfirina. Otras dos vías, la de interconversión de pentosa y glucuronato y la de biosíntesis de monobactámicos, tuvieron baja significancia por lo que se descartaron de los análisis posteriores (Figura 3.5B). En el caso del SR, el enriquecimiento sólo se encontró en la biosíntesis de antocianinas, de monobactámicos y en la interconversión de pentosas y glucuronato, mientras que SN la mayoría de las rutas estuvieron sobrerrepresentadas (Figura Suplementaria 3.5).

Estas vías estuvieron relacionadas con la fotosíntesis (proteínas antena y metabolismo de porfirina), con la generación de monosacáridos y poder reductor (ruta de las pentosas fosfato), biosíntesis y metabolismo de metabolitos secundarios y metabolismo de nitrógeno y aminoácidos (antocianinas, carotenoides, esqueleto de terpenoides, arginina, alanina, aspartato y glutamato, ácido 2-oxocarboxílico) y funciones celulares relacionadas con la degradación de proteínas (proteasoma) y el ritmo circadiano (Figura 3.5B).

Con respecto a la biosíntesis de capsaicinoides, el metabolismo del ácido 2-oxocarboxílico se relaciona con los precursores valina y leucina. Sin embargo, no se encontró enriquecimiento en una ruta más directa con la biosíntesis de capsaicinoides, es decir, en la ruta de los fenilpropanoides o en la síntesis de ácidos grasos. Se decidió entonces analizar la expresión de genes individuales de esta vía por tipo de suelo, así como el comportamiento de genes regulados de las rutas metabólicas enriquecidas para revisar su comportamiento de manera individual (comprobar si los genes están regulados individualmente por el tipo de suelo o sólo el comportamiento como grupo de genes de las vías se vio afectado).



Rutas enriquecidas

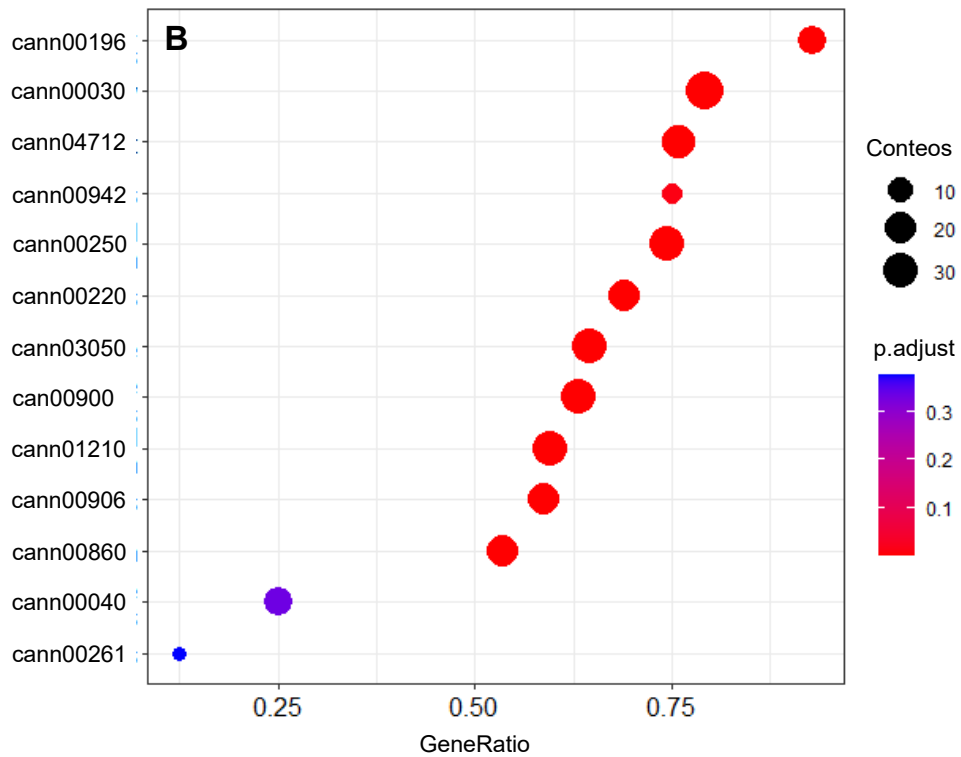


Figura 3.5. Análisis global de expresión diferencial (DGE) y rutas metabólicas enriquecidas (GSEA) en los transcriptomas de frutos de plantas de chile habanero crecidas en suelo negro (SN) y suelo rojo (SR). A) Gráfica de volcán en la comparación de la DGE del transcriptoma de los frutos de chile habanero. Se relacionan la magnitud de la comparación de SR contra SN como LFC ($\log_2\text{FoldChange}$) y la confianza de la comparación (pvalue). En total se representan 92867 genes. Los puntos azules representan aquellos genes con mayor expresión en SR ($\text{LFC} \geq 1.5$), o en SN ($\text{LFC} \leq -1.5$) con $\text{pvalue} < 0.00835$ (correspondiente a $\text{padj} < 0.1$). Con puntos rojos se muestran aquellos genes sin expresión diferencial en base al criterio de LFC y pvalue. Los genes 'outliers' seleccionados se muestran encerrados en círculos. La clave corresponde a la lista de genes en la Tabla Suplementaria 3.1A y 3.1B. B) Análisis de enriquecimiento de rutas GSEA por términos KEGG de los genes del transcriptoma comparando SR contra SN. El tamaño de los círculos representa el número de genes asociados con la ruta enriquecida en un tipo de suelo (conteos), y su color representa la confianza en la comparación (pajust), mientras que el eje "x" representa la proporción de genes anotados en esa ruta en un suelo contra el total de genes anotados en la ruta (geneRatio.). El eje "y" muestra las rutas metabólicas enriquecidas por su código KEGG. cann00196: proteínas antena de fotosíntesis, cann00030: ruta de las pentosas fosfato, cann04712: ritmo circadiano, cann00942: biosíntesis de antocianinas, cann00250: metabolismo de alanina, aspartato y glutamato, cann00220: biosíntesis de arginina, cann03050: proteasoma, can00900: biosíntesis de terpenoides (backbone), cann01210: metabolismo del ácido 2-oxocarboxílico, cann00906: biosíntesis de carotenoides, cann00860: metabolismo de porfirina, cann00040: interconversión pentosa y glucuronato, cann00261: biosíntesis de monobactámicos. El análisis DGE se realizó con DESeq2 y el GSEA con ClusterProfiler v4.4.4

3.3.4. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides

Para el análisis de expresión se realizó la búsqueda de genes de la ruta de síntesis de capsaicinoides, así como de al menos 2 genes regulados de cada ruta metabólica enriquecida del análisis de la Figura 3.5B. Esta se realizó mediante un blastn local por bioedit contra el transcriptoma ensamblado. En la Tabla 3.3 y la Tabla Suplementaria 3.2 se muestran los resultados de la búsqueda de genes y de su anotación contra la base de datos del NCBI o del KEGG.

Tabla 3.3. Lista de *contigs* usados para el análisis de expresión de rutas metabólicas enriquecidas y genes de la ruta de síntesis de capsaicinoides en el transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro o suelo rojo. Se usaron genes homólogos de la ruta de síntesis de capsaicinoides y de las vías metabólicas enriquecidas de la Figura 3.5B para realizar la búsqueda por blastn local contra la biblioteca de unigenes. Los genes están agrupados de acuerdo a los clados formados en la Figura 3.6

Clado	Contig	Descripción del gen	Clave	Gen de referencia (ID)*
1	c97246_g1_i1	Proteína 2 de unión a clorofila del complejo LHC I	LHCA2**	cann107838850
	c66898_g1_i1	Cetol-ácido reductoisomerasa	KARI**	cann107856177
	c51789_g1_i1	Proteína 3 de unión a clorofila del complejo LHC II	LHCB3**	cann:107878708
2	c97797_g1_i1	Glutamato-1-semialdehído 2,1-aminomutasa	GSA**	cann107863035
	c107703_g1_i1	Ornitina transcarbamilasa	OTC	cann107866980
	c49639_g1_i1	Regulador de pseudorespuesta 7	PRR37	cann107850219
	c80334_g1_i1	15-cis fitoeno desaturasa	PDS	cann107861625
	c63879_g4_i1	Subunidad N2 regulatoria del proteasoma 26S	RPN2	cann107854708
	c37467_g1_i1	Glutamato sintasa	GOGAT	cann107855963
	c66114_g8_i1	Gamma-aminobutirato transaminasa 1	GABAT1	cann107862720
	c98787_g1_i1	Glutamil-tRNA reductasa	GluTR**	cann107867325
	c10711_g1_i1	1-Desoxi-xilulosa-5-fosfato sintasa	DXS**	cann107852315
	c73194_g1_i1	Arginasa 1	ARG1	cann107873005
	c62067_g3_i1	Glutamina sintetasa	GS	cann107870519
c24104_g1_i1	Aspartato cinasa	AK	cann107838766	
c2734_g1_i1	6-fosfogluconato deshidrogenasa	6PGD***	cann107850704	
3	c61017_g1_i2	UDP-glucosiltransferasa	UGTII	cann107870274, cann107867844
	c62154_g1_i3	15-cis fitoeno sintasa 2	PSY2	cann107859651
	c65913_g4_i6	Factor de transcripción MYB	MYB	MK074771.1, MK074772.1, MF062086.1
	c61017_g2_i1	UDP-glucosiltransferasa	UGTI	cann107870274
	c54416_g1_i1	Hipocótilo alargado tardío	LHY	cann107845265
	c64923_g1_i1	Homólogo de <i>PUN1</i>	AT3D2	MCIT02000002.1 (PHU26685.1,)
	c66114_g6_i1	Gamma-aminobutirato transaminasa 3	GABAT3	cann107841786
c72621_g1_i1	1-Desoxi-xilulosa-5-fosfato sintasa 2	DXS2**	cann107850768	
c10708_g1_i1	Glucosa-6-fosfato deshidrogenasa	G6PDH***	cann107860735	

CAPÍTULO III

4	c67203_g1_i1	Subunidad T4 regulatoria del proteasoma 26S	RPT4***	cann107840436
	c65744_g2_i1	Cetoacil-ACP sintasa	KAS	AF085148.1
	c89513_g1_i1	Acil-ACP tioesterasa	FAT	AF318288.1
	c97242_g1_i1	Acil-ACP reductasa	KR	EU616561.1
	c65172_g1_i1	Capsaicina sintasa, <i>PUN1</i>	AT3D1	MCIT02000002.1 (PHU26684.1)
5	c64881_g10_i1	Aminotransferasa putativa (de la vanillilamina)	pAMT	AF085149.1

*Número de accesoión del gen de referencia en la base de datos del NCBI o del KEGG

**Diferencialmente expresados de acuerdo al análisis por DESeq2 (padj.<0.05)

***Diferencialmente expresados de acuerdo al análisis por DESeq2 (padj.<0.1)

En la Figura 3.6 y la figura suplementaria 3.6 se muestra el análisis de expresión de los genes de las rutas metabólicas mencionadas en los cuatro transcriptomas (SN1, SN2, SR1 y SR2). Por sus patrones y niveles de expresión, los genes se agruparon en cuatro clados principales. El primero de éstos incluyó los 3 genes altamente expresados, con 450-2000 TPM (*LHCA2*, *LHCB3* y *KARI*) (Tabla 3.3 y Tabla Suplementaria 3.2 y). Dos de estos corresponden a genes del complejo de proteínas-clorofila antena de la fotosíntesis (*LHCA2* y *LHCB3*). El primero fue homólogo a proteínas 2 del complejo de recolección de luz del fotosistema I (*LHCA2*), mientras que el segundo fue homólogo a proteínas 3 del complejo de recolección de luz del fotosistema II (*LHCB3*). Además, estos genes fueron más expresados en SN (padj. <0.05), indicando que la expresión de genes de los complejos antena de unión a clorofila de fotosíntesis fue afectada por el tipo de suelo, lo que también fue sugerido por el análisis de rutas metabólicas enriquecidas, donde esta vía fue la más sobrerrepresentada al contrastar tipos de suelo (Figura 3.5B).

En este clado también se encontró un gen altamente expresado del metabolismo del ácido 2-oxocarboxílico y de la valina, leucina e isoleucina, la cetol-ácido reductoisomerasa (*KARI*) (Tabla 3.3). Este fue más expresado en SN (padj. <0.05), aunque su contraste fue menor que en los genes de las proteína-antena (Figura 3.6 y Figura Suplementaria 3.6A).

El segundo clado incluyó 13 genes con diferentes niveles de expresión (Figura 3.6 y Figura Suplementaria 3.6B), pero menores a los del clado I (38 a 166 TPM), y de variadas rutas metabólicas. En primer lugar, se tiene a los genes *GSA* (glutamato-1-semialdehído 2,1

aminomutasa, del metabolismo de la porfirina), *OTC* (ornitina transcarbamilasa, del metabolismo de la arginina), *PRR37* (regulador de pseudorrespuesta 7 del ciclo circadiano) y *PDS* (15-cis fitoeno desaturasa, de la síntesis de carotenoides) (Tabla 3.3).

En *GSA* se encontró una mayor expresión en SN ($p_{adj} < 0.05$), mientras que en *PDS*, *OTC* y *PRR37* no hubo diferencia entre tipos de suelo, aunque SN1 tuvo tendencia a una mayor expresión (Figura Suplementaria 3.6B). En otro subclado se tuvieron a *RPN2* (una subunidad regulatoria del proteasoma), *GOGAT* (glutamato sintasa) y *GABAT1* (gamma-aminobutirato transmaninsa 1) (Tabla 3.3).

No hubo diferencia en la expresión de los tres genes entre tipo de suelo. Se observó una tendencia similar a la del subclado anterior, donde, independientemente de la diferencia entre suelos, SN1 tuvo una expresión mayor, con excepción de *GABAT1*, donde en SR1 hubo la más alta. La mayoría de estos genes, a excepción de *PRR37* y *PDS*, tienen en común que están relacionados o participan en el metabolismo de aminoácidos (glutamato) y del nitrógeno. Además, tuvieron una expresión similar (38 a 72 TPM) (Tabla Suplementaria 3.2). Esto podría sugerir una relación en la regulación de su expresión, aun cuando la mayoría no tuvo diferencia entre tipo de suelo.

Otro de los subclados tuvo a *GluTR* (glutamil-tRNA reductasa, del metabolismo de porfirina) y a *DXS* (1-desoxi-xilulosa-5-fosfato sintasa, de la síntesis del esqueleto de terpenoides), en donde hubo una mayor expresión en SN ($p_{adj} < 0.05$).

La *ARG1* (arginasa 1) no tuvo diferencia entre tipos de suelo (Figura 3.6 y Figura Suplementaria 3.6B). Finalmente, la *GS* (glutamina sintetasa), *AK* (aspartato cinasa) y *6PGD* (6-fosfogluconato deshidrogenasa, de la ruta de las pentosas fosfato) tuvieron una expresión alta en las cuatro muestras, especialmente *6PGD* (126 a 166 TPM). Sólo *6PGD* tuvo diferencia entre tipos de suelo ($p_{adj} < 0.1$) (Figura 3.6 y Tabla 3.3). Además, la alta expresión de los 3 genes refleja su importancia en sus rutas metabólicas (metabolismo del glutamato, aspartato y alanina, del ácido 2-oxocarboxílico; y la ruta de las pentosas fosfato, respectivamente).

Teniendo en cuenta los genes por ruta metabólica, se destaca que *GluTR* y *GSA*, ambos del metabolismo de porfirina, fueran más expresados en SN (Tabla 3,3 y Figura 3.6) y que esta vía estuvo enriquecida al comparar tipos de suelo (Figura 3.5B), sugiere su importancia en frutos cultivados en diferentes suelos. *6PGD* también fue más expresado en SN. La ruta de las pentosas fosfato fue también de las enriquecidas entre tipo de suelo (Figura 3.5B), lo que

refleja la importancia de esta ruta que está relacionada con la obtención de pentosas y NADPH para procesos de síntesis.

DXS, también más expresado en SN, sugirió la importancia del tipo de suelo para la ruta metabólica enriquecida de síntesis del esqueleto de terpenoides. La expresión de *PDS*, de la síntesis de carotenoides, no fue afectada por el tipo de suelo (Figura 3.5B, Figura 3.6 y Tabla 3.3). Aunque otros genes no hayan sido más expresados entre tipos de suelo, varios genes del metabolismo del nitrógeno y de aminoácidos estuvieron en este clado. Estas rutas y genes fueron la biosíntesis de arginina (*OTC* y *ARG1*), metabolismo del aspartato, glutamato y alanina (*GABAT1*, *GS* y *GOGAT*), metabolismo del ácido 2-oxocarboxílico (*AK*), además *RPN2* (proteasoma), relacionado con el catabolismo de proteínas, también se encontró en este clado. Esto sugiere la regulación conjunta de estos genes y las rutas metabólicas relacionadas con los aminoácidos y el nitrógeno.

El tercer clado tuvo 8 genes, de baja expresión y pertenecientes a varias rutas metabólicas (0.7 a 43 TPM) (Figura 3.6, Tabla 3.3 y Tabla Suplementaria 3.2). Los genes fueron *UGT1* y *UGTII* (UDP-glucosil transferasa), relacionados con la biosíntesis de antocianinas; *PSY2* (15-cis fitoeno sintasa 2), de la biosíntesis de carotenoides; *MYB* (factor de transcripción relacionado con la síntesis de capsaicinoides); *LHY* (hipocótilo alargado tardío) del ciclo circadiano; *AT3-D2* (aciltransferasa relacionada con la CS); *GABAT3* (gamma-aminobutirato transaminasa 3) relacionado con el metabolismo de glutamato, aspartato y alanina; y *DXS2* (1-desoxi-xilulosa-5-fosfato sintasa 2), de la síntesis del esqueleto de los terpenoides. La mayoría de los genes de este clado no tuvieron diferencia en su expresión por tipo de suelo, incluyendo los genes representantes de la ruta de síntesis de antocianinas (*UGT*) que no tuvieron una expresión diferente por tipo de suelo, aun cuando esta ruta estuvo enriquecida al comparar tipos de suelo (Figura 3.5B), aunque los genes *PSY2* y *UGTII* tuvieron cierta tendencia a una mayor expresión en SN1. Por el contrario, *DXS2* tuvo mayor expresión en SN ($p_{adj} < 0.05$) (Figura 3.6 y Figura Suplementaria 3.6B). Además, *GABAT3* y *DXS2* tuvieron una expresión relativamente más alta que los otros genes en este clado (9.3 a 42.7 TPM), que sugiere una regulación distinta de éstos (Figura 3.6 y Figura Suplementaria 3.6B).

Varios de los genes de este clado fueron de rutas presentes también en genes del clado 2. Como varios genes del clado 2 y 3 tuvieron diferentes niveles de expresión, podría indicar distintos niveles de regulación para genes en la misma ruta metabólica. Esto se observó en parejas como *PSY2* con *PDS*, *LHY* con *PRR37*, *GABAT1* con *GABAT3*, y *DXS* con *DXS2*.

Respecto a éstos últimos, al estar relacionados con el metabolismo de terpenoides y ser más expresados en SN, esto muestra la importancia del tipo de suelo en la síntesis de estos metabolitos (Figura 3.6 y Figura Suplementaria 3.6B).

En los clados 4 y 5 hubo 6 genes y 1 gen, respectivamente. Los niveles de expresión de estos genes fueron, en general, mayores a los de los clados 2 y 3, pero menores a los del clado 1 (77 a 696 TPM) (Tabla Suplementaria 3.2). En su mayoría pertenecieron a la ruta de biosíntesis de capsaicinoides (*KAS*, *FAT*, *KR*, *AT3-D1* y *pAMT*). Además, hubo un gen relacionado al proteasoma (*RPT4*) y uno relacionado a la ruta de las pentosas fosfato (*G6PDH*) (Tabla 3.3 y Figura 3.6). Se encontró en estos dos últimos una mayor expresión en SN ($p_{adj} < 0.1$). *FAT* y *KAS*, ambos relacionados con la síntesis de ácidos grasos, se agruparon en otro subclado. Los genes *KR* y *AT3-D1* tuvieron la expresión más alta en el clado 4 (192 a 379 TPM).

Este comportamiento es esperado en el gen de la CS, *AT3-D1*, en frutos de 25 DPA, ya que se trata de un gen que determina la pungencia en los frutos y que en etapas intermedias de desarrollo suele tener expresiones altas. El gen *pAMT* tuvo la expresión más alta entre los genes de la síntesis de capsaicinoides (338 a 696 TPM) (Figura 3.6 y Figura Suplementaria 3.6C), pero se mantuvo en un clado diferente del resto de éstos. Su alta expresión también fue esperada por razones similares a las de *AT3*, tratándose ambos de genes clave en la ruta de síntesis de capsaicinoides. Su ubicación en el clado 5 puede sugerir un nivel de regulación separado de los otros genes de la ruta de síntesis de capsaicinoides analizados.

Varios de los genes del clado 4 pertenecieron a rutas donde también hubo genes de otros clados y con otros niveles de expresión. Entre *RPT4* y *RPN2*, ambos relacionados con el proteasoma, *RPT4* no sólo tuvo mayor expresión, sino que fue más expresado en SN. Esto indica una posible regulación diferente y diferentes funciones de ambos genes, pese a pertenecer a la misma vía, ya que son genes de distintas subunidades del proteasoma (subunidad T4 y subunidad N2, respectivamente). También, una expresión diferencial en genes de esta ruta refuerza el análisis de enriquecimiento de rutas metabólicas, donde el proteasoma fue una de las más sobrerrepresentadas al comparar suelos (Figura 3.5B).

Por otro lado, los genes de la ruta de las pentosas fosfato *G6PDH* y *6PDG*, ambos estuvieron más expresados en SN, además que fue una de las rutas más enriquecidas entre suelos (Figura 3.5B). Ambos genes estuvieron ubicados en distintos clados, sugiriendo distinto nivel de regulación. En los genes de la ruta de biosíntesis de capsaicinoides. *AT3-D1* y *AT3-D2* se

ubicaron en distintos clados, siendo *AT3-D1* mucho más expresado. Al comparar la CDS y la secuencia traducida de los ortólogos de *AT3-D1* y *AT3-D2* se encontró que tuvieron 89.63% y 86.57% de identidad, respectivamente. Además, *AT3-D1* se ubicó con otros ortólogos de *AT3-D1* en *C. chinense*, mientras que *AT3-D2* se ubicó con otro ortólogo de *AT3-D2* (Figura Suplementaria 3.7), por lo que se tratan de copias distintas del mismo gen, lo que se ha reportado anteriormente en el genoma de *Capsicum* (Qin *et al.*, 2014). Más esfuerzos deben hacerse para identificar el papel más específico de los diferentes ortólogos *AT3-D1* y *AT3-D2* durante el desarrollo del fruto.

Con respecto a los niveles de capsaicinoides en los frutos (Figura 3.3) y la expresión de los genes de la ruta de síntesis de estos metabolitos (Figura 3.6), aunque a 25 DPA sí se observó mayor contenido de capsaicina en SN, el nivel de expresión de los genes relacionados con la ruta de síntesis no reflejó esa diferencia al no tener distinta expresión; aunque, el nivel alto de la expresión de *AT3-D1* y *pAMT* refleja la importancia de estos metabolitos en los frutos de chile habanero. De esto se concluye, que, al menos a 25 DPA, el efecto del suelo en la expresión de los genes de la síntesis de capsaicinoides no se va a observar, pero sí se observa en el fenotipo de los frutos y probablemente los niveles de expresión se deban analizar en frutos en otros estados de desarrollo antes de 25 DPA, especialmente en SN, donde los frutos alcanzaron un nivel de capsaicinoides que conserva a 45 DPA posiblemente siendo un límite máximo en la acumulación.

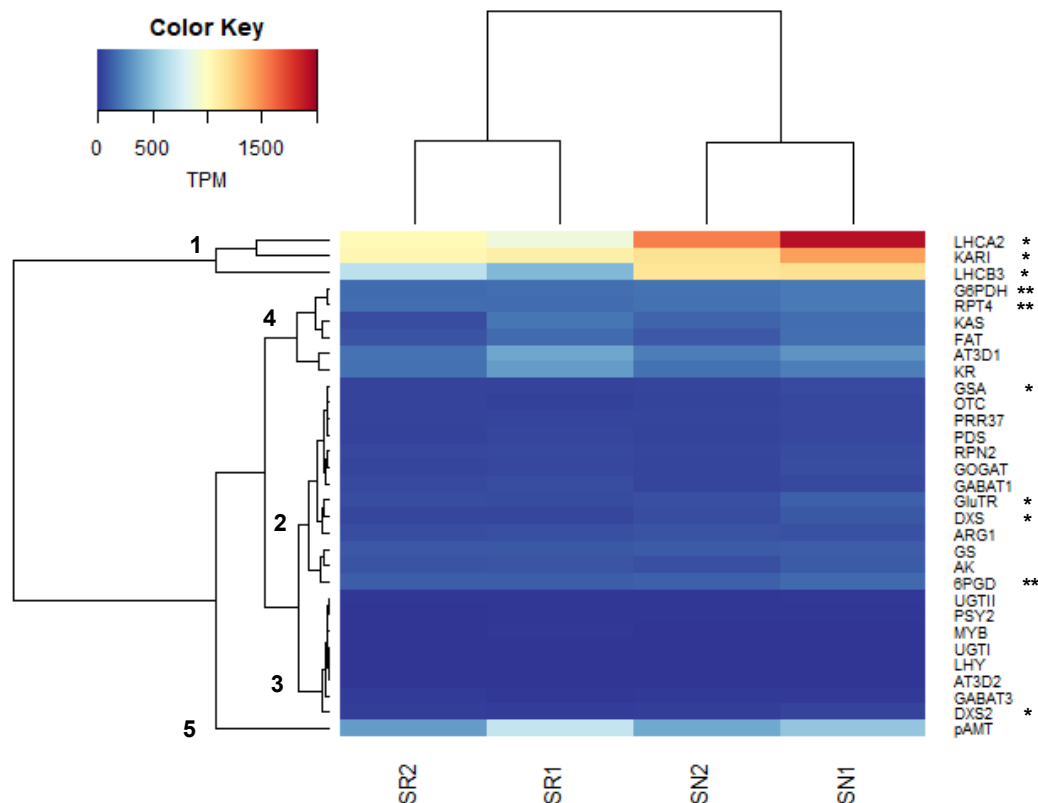


Figura 3.6. Análisis de la expresión de genes seleccionados de las rutas enriquecidas de la Figura 3.5B y de la ruta de síntesis de capsaicinoides. La expresión se muestra en transcritos por millón (TPM) en el transcriptoma de las dos réplicas biológicas de frutos en suelo negro (SN1 y SN2) y suelo rojo (SR1 y SR2). Se muestra la numeración de los distintos clados. La clave de los genes corresponde al código de la Tabla 3.3. La gráfica se realizó en R con la herramienta gplots. El código de los genes corresponde a la lista de la Tabla 3.3. *Genes diferencialmente expresados entre tipos de suelo ($p_{adj} < 0.05$). **Genes diferencialmente expresados entre tipos de suelo ($p_{adj} < 0.1$), de acuerdo al análisis de DESeq2

3.4. DISCUSIÓN

3.4.1. Producción de flores y frutos y caracterización de frutos en plantas de chile habanero crecidas en dos tipos de suelo

En este trabajo, se reportó el efecto de dos tipos de suelo de Yucatán con características contrastantes sobre la producción de frutos y de capsaicinoides en chile habanero. Los resultados de producción de frutos mostraron un aborto floral relativamente alto (85%) en ambos tipos de suelo. La producción total para la variedad Jaguar fue de en promedio 35 a 40 frutos y entre 31 a 33 frutos de 45 DPA, que son frutos que ya están llegando a estado de

madurez completa. Ramírez-Meraz *et al.* (2018) reportan, en general para cielo abierto y sistemas de agricultura protegida, que esta variedad de chile habanero tiene frutos de 6.5 a 10 g y un largo de 3.8 a 5.5 cm. En el presente trabajo los frutos más desarrollados, mayores a 4 cm de largo, tuvieron valores semejantes, con 6.95 a 6.96 g de peso promedio.

Al comparar los frutos por rango, en SR se obtuvo un mayor número total de frutos más grandes, mientras que en SN hubo un mayor número de frutos de rango de tamaño inferior (2-2.99 cm). Este resultado fue similar al obtenido por Medina-Lara *et al.* (2019) también en plantas de la variedad Jaguar en SR y SN. Sin embargo, estos autores usaron agua de alta conductividad y obtuvieron un mayor contraste en las diferencias en el tamaño de los frutos por tipo de suelo, con una tendencia a un mayor número de frutos más pequeños en SN, incluso menores a 1 cm de largo. Esto sugiere que con riego con agua de baja conductividad eléctrica en SR se favorece la obtención de frutos con mayor desarrollo, pero en SN todavía se pueden obtener frutos desarrollados (3-3.99 cm) por lo que la diferencia global en la producción total de frutos, aun dividiéndolos por tamaño, puede considerarse pequeña entre tipos de suelo.

Los suelos negros tienen un mayor contenido de los nutrientes que los suelos rojos (Estrada-Medina *et al.*, 2016; Bautista-Zúñiga *et al.*, 2003). Aun así, es importante también considerar la movilidad de estos nutrientes de la matriz del suelo a la solución del suelo y hacia las raíces de las plantas (Borges-Gómez *et al.*, 2005). El aprovechamiento más eficiente de los nutrientes en plantas de chile habanero se ha observado anteriormente en SR en comparación a SN, pese a su menor composición nutricional (Medina-Lara *et al.*, 2019). Sin embargo, esto ha sido bajo riego con agua de alta conductividad eléctrica, contrario a lo realizado en el presente trabajo, por lo que se tendría que estudiar el aprovechamiento de los nutrientes en chile habanero con el tipo de agua usada.

Las diferencias de textura (composición de arena, limo y arcilla) y contenido de materia orgánica entre los suelos negro y rojo también son importantes para la movilidad y aprovechamiento de nutrientes, como ya se ha mencionado (Medina-Lara *et al.*, 2019; Bautista *et al.*, 2005). Esta diferencia en composición influye en la estructura (agregados de partículas del suelo), la porosidad y la densidad aparente del suelo, que afectan la filtración del agua y la retención de componentes de la matriz del suelo (Schaeztl y Anderson, 2005). Los suelos rojos, con mayor densidad aparente (Bautista *et al.*, 2005), tienen menor porosidad. Los suelos negros, con mayor contenido de materia orgánica, podrían tener mayor proporción de poros grandes, que filtren la solución de suelo más que en suelos rojos, que afecta la difusión de los

nutrientes del suelo hacia las raíces de las plantas, aunque también favorece un suelo estructurado al cohesionar las partículas del suelo. Por otro lado, suelos con alto contenido de partículas de tamaño pequeño con un área superficial grande, como los suelos rojos, retienen más fuertemente los componentes del suelo, lo que afecta la movilidad de los componentes del suelo (Schaeztl y Anderson, 2005). También se debe de considerar la forma química en la que se encuentran los nutrientes en el suelo. En los suelos negros se ha reportado un mayor contenido de nitrato que en suelo rojo (Bautista-Zúñiga *et al.*, 2003), que es una forma química de nitrógeno que el chile habanero puede asimilar. Otros nutrientes, como el fósforo, si se toma en cuenta que los valores de pH reportados en la literatura suelos negro y rojo son de neutros a ligeramente alcalinos (Medina-Lara *et al.*, 2019; Estrada-Medina *et al.*, 2016), estaría en baja solubilidad y poco disponible para la planta (Borges-Gómez *et al.*, 2014). Además, las diferentes composiciones en poblaciones microbianas (bacterias y hongos) entre los suelos negro y suelo rojo también pueden influir en la forma química en la que se pudieran encontrar los nutrientes y la materia orgánica del suelo. Por estos motivos, sería importante determinar las formas químicas disponibles de éstos para tener un mejor acercamiento del aprovechamiento de la planta hacia los componentes del suelo.

3.4.2. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en dos tipos de suelo

Los capsaicinoides son considerados entre los metabolitos más deseables en la calidad del chile habanero por la sensación de picor que les dan a los frutos, lo que es buscado por muchos consumidores. Por esta característica, y adicionalmente por las propiedades bioactivas que poseen, son importantes a nivel industrial (Wang *et al.*, 2022) Por ello fue fundamental determinar su el efecto del tipo de suelo sobre el contenido de capsaicinoides en los frutos, y conocer que suelo puede convenir si lo que se busca es tener chiles con un mayor contenido de capsaicinoides. Los resultados mostraron un aumento en el contenido de capsaicinoides de 25 a 45 DPA y luego disminución a 60 DPA en SR, pero en SN desde los 25 DPA hubo un máximo en el contenido de capsaicinoides que bajo sólo hasta los 60 DPA. El máximo contenido de capsaicinoides fue de 11.30 mg/g PS y 11.48 mg/g de PS de fruto, o 169500 y 172200 SHU (unidades de Scoville) para SN y SR, respectivamente (Tabla Suplementaria 3.3). Este grado de picor fue superiores al mostrado en otras especies domésticas de *Capsicum* (*C. annum*, *C. frutescens*, *C. pubescens* y *C. baccatum*) (Sun *et al.*, 2022), siendo *C. chinense* en general considerada la especie de chile más picante, aunque fue inferior al obtenido en otras variedades de chile habanero cultivadas en Yucatán (Canto-Flick *et al.*, 2008). También fueron

similares al contenido de capsaicinoides obtenido por Medina-Lara *et al.* (2008) para chile habanero en suelo rojo, donde se encontró que, bajo fertilización con urea, no fue posible aumentar el contenido de capsaicinoides tras llegar a un límite de acumulación, no diferenciándose del tratamiento sin fertilización (Medina-Lara *et al.*, 2008). Comparando dichos resultados con este trabajo, se sugiere que ambos suelos son buenos para tener un máximo de contenido de capsaicinoides en frutos de 45 DPA, que están alcanzando la madurez completa, llegando a un posible máximo o umbral sin necesidad de aplicar fertilización de nitrógeno. Se ha reportado que es posible aumentar el contenido de capsaicinoides bajo condiciones de estrés u otros regímenes de fertilización como la fertilización orgánica en suelos con diferentes características con respecto a tratamientos sin fertilización, como reportaron Das *et al.* (2016) en *C. chinense* 'Borbhut' en dos diferentes suelos de regiones de la India con diferentes contenidos de textura, nitrógeno, fósforo, potasio y materia orgánica similar al contraste entre suelos negro y rojo.

Con respecto al comportamiento del contenido de capsaicinoides durante el desarrollo del fruto, los resultados sugieren una influencia en el tipo del suelo sobre la síntesis y degradación de los capsaicinoides, con efecto diferente en distintas etapas de desarrollo. Considerando el contenido de capsaicinoides que se observa en un determinado punto es una el reflejo de su síntesis y degradación, entonces entre 25 DPA a 45 DPA en SR predominó la síntesis, aumentando su contenido de capsaicinoides, y de 45 a 60 DPA predominó la degradación de capsaicinoides, disminuyendo. En SN de 25 a 45 DPA ni la síntesis o degradación parecieron predominar, y el proceso de síntesis debió ocurrir en etapas más tempranas, por lo que después de 25 DPA el metabolismo de capsaicinoides no modificó el contenido total de éstos hasta los 60 DPA, donde predominó la degradación, disminuyendo su contenido. Las características diferentes de los suelos afectan por lo tanto el metabolismo de los capsaicinoides. La expresión de genes como CS es sensible al tipo de suelo (Das *et al.*, 2021). En los suelos usados, una posibilidad es que el contenido de nitrógeno, su disponibilidad (Medina-Lara *et al.*, 2019) o la forma química en la que se encuentre éste afecte la expresión de los genes de la ruta de síntesis de capsaicinoides o la actividad de sus enzimas (Zhang *et al.*, 2020a). En este sentido, sabe que el contenido de nitrato en frutos y placenta de chile habanero afecta su acumulación de capsaicinoides (Monforte-González *et al.*, 2010). El diferente contenido de materia orgánica y de ácidos fúlvicos y húmicos de en suelos distintos también tiene relación con el contenido de capsaicinoides en frutos (Das *et al.*, 2016) y el diferente contenido de materia orgánica que hay entre suelo negro y suelo rojo probablemente

sea otro factor a considerar para los cambios en la acumulación de capsaicinoides (Medina-Lara *et al.*, 2019). De acuerdo a los resultados obtenidos en este trabajo y teniendo en cuenta las diferentes propiedades entre los suelos, en SN se estaría estimulando el metabolismo de capsaicinoides de tal modo que la acumulación alcanza un máximo en etapas intermedias de desarrollo del fruto (25 DPA) mucho antes que en las plantas crecidas en SR. Por lo anterior, se seleccionó esta etapa de desarrollo para el análisis transcriptómico en las plantas crecidas en los diferentes tipos de suelo.

3.4.3. Análisis general del transcriptoma y de rutas metabólicas enriquecidas en frutos de plantas de chile habanero crecidas en dos tipos de suelo

3.4.3.1. Análisis general del transcriptoma

Los frutos de 25 DPA se analizaron por medio de transcriptómica para analizar la expresión de genes que se relacionan con la síntesis de capsaicinoides, ya que las etapas intermedias de desarrollo de frutos de *Capsicum* son activas en la expresión hacia arriba de genes involucrados con la síntesis de capsaicinoides en *Capsicum* (Zhang *et al.*, 2016; Martínez-López *et al.*, 2014).

Los resultados de la calidad mostraron que, tras el filtrado y recorte de secuencias de baja calidad, la mayor parte de los *reads* tuvieron altos valores phred, por encima de Q20 (más del 98% de las bases) y de Q30 (más del 95% de las bases), con una calidad promedio de read Q36. El número de *reads* tras el recorte representó entre el 98 al 99% de los *reads* originales. El valor phred es un parámetro de la precisión usado en genomas y transcriptomas que indica una probabilidad del error de secuenciación (encontrar una base incorrecta o mal secuenciada), y mientras más bajo sea mayor es la probabilidad del error, equivaliendo Q20 a un 1% de error y Q30 a 0.1% de error (Illumina, 2011). Los resultados de la calidad mostraron, por tanto, que los valores de precisión fueron altos, un indicativo de un buen transcriptoma para su uso. El ensamblado del transcriptoma tuvo 99159 unigenes en 68.37 Mb, con un N50 de 1143 pb, un promedio de longitud de 689.55 pb, y *contigs* de 201 a 9421 pb. Nuestro ensamblado es comparable al obtenido en otros transcriptomas de *Capsicum* tanto en flores como en frutos de distinto estado de madurez (Chhapekar *et al.* 2020; Liu *et al.*, 2013). En el remapeo de *reads*, se obtuvieron altos porcentajes de fragmentos mapeados contra la biblioteca de unigenes (78-82%). El remapeo de *reads* contra un transcriptoma obtenido por ensamblado de *novo* es importante como indicador de su calidad, y generalmente se recomienda un valor de

aproximadamente 80% o superior, lo que se obtuvo en el presente trabajo, indicando que la mayoría de los *reads* fueron usados para la construcción de la biblioteca de unigenes (Raghavan *et al.*, 2022), y es un requisito para la estimación de la abundancia de los transcritos y de los cálculos de expresión, por medio de los conteos de los fragmentos alineados a los *contigs* (Haas *et al.*, 2013). Por lo tanto, el remapeo indicó que una gran cantidad de información de la secuenciación se conserva en el transcriptoma ensamblado para usarse en los análisis de anotación funcional y expresión.

3.4.3.2. Análisis de genes diferencialmente expresados y de rutas metabólicas enriquecidas entre los frutos de chile habanero en suelo negro y suelo rojo

Los genes de cada biblioteca (SN1, SN2, SR1 y SR2) se anotaron con la base de datos GO. Esto permite agruparlos dentro de tres categorías funcionales: BP, CC y MF (Gene Ontology Consortium, 2004). No se encontraron diferencias en la proporción de genes dentro de las subcategorías pertenecientes a los tres términos principales, esto sugiere una funcionalidad similar en los frutos de chile habanero de 25 DPA independientemente del tipo de suelo, dentro de las categorías GO principales.

Al analizar la expresión diferencial de genes entre tipos de suelo, en SN destacó que varios de los genes más expresados ($LFC < -1.5$, $pvalue < 0.00835$) estuvieron relacionados con la fotosíntesis. Estos fueron ortólogos de proteínas de unión a clorofila de los complejos de recolección de luz (LHC), y de RuBisCO. Se sabe que la expresión de los genes relacionados con la fotosíntesis está controlada por el estado de desarrollo del fruto, disminuyendo su expresión en frutos entrando a madurez y con el cambio de coloración (Borovsky *et al.*, 2019; Powell *et al.*, 2012), aunque los frutos de 25 DPA no mostraron una diferencia en el color de su fenotipo por el tiempo de desarrollo, a los 45 DPA, donde los frutos ya estaban en etapa de cambio de coloración o completamente maduros (naranjas), hubo mayor proporción de frutos en estado de cambio de coloración y menor proporción de frutos aún en color verde en SN en comparación a SR (Tabla Suplementaria 3.4). Es posible también que en los suelos negros se estimule más el aparato fotosintético en frutos de 25 DPA que en los suelos rojos, lo cual es favorable en la maduración posterior desde un punto de vista productivo para obtener frutos con mayor contenido de azúcares y metabolitos secundarios (Lupi *et al.*, 2019; Nguyen *et al.*, 2014). Otros ortólogos encontrados fueron de proteínas de dedo de anillo-H2, que están relacionados con crecimiento, desarrollo y la respuesta a factores y estrés abióticos (Han *et al.*, 2022) y con el desarrollo de semillas (Xu y Li, 2003); de BREVIS RADIX, relacionada con el

desarrollo del fruto y del número de lóculos en *C. annuum* (Ma *et al.*, 2022); de albúmina, que se han encontrado en semillas de *C. annuum* y que sirven de almacenamiento, y tienen propiedades antimicrobianas (Ribeiro *et al.*, 2012). En SR, entre los genes más expresados ((LFC>1.5, pvalue<0.00835) hubo un ortólogo de la 3-cetoacil-CoA sintasa (KCS), relacionada con la síntesis de ácidos grasos. Otro gen encontrado fue un ortólogo de la calmodulina lisina N-metiltransferasa, donde la calmodulina está relacionada con la señalización por calcio entre la pared celular y el citoplasma, y en el crecimiento del tubo polínico (Gong *et al.*, 2022; Hepler y Winship, 2010). Otros genes con mayor expresión en SR fueron homólogos de la proteína mago nashi, relacionada con la fertilidad (Ihsan *et al.*, 2015); de proteoglucano rico en prolina, extensinas; proteínas F-box; y poliubiquitina y proteína de dominio BURP, que participan en procesos como la respuesta a estímulos ambientales, el desarrollo floral, el desarrollo de los frutos y sus paredes celulares, incluyendo las etapas de maduración, ablandamiento y abscisión (Leszczuk *et al.*, 2020; Zhang *et al.*, 2019; Park *et al.*, 2015b; Sommer-Knudsen *et al.*, 1996). Esto diferencia a los genes más expresados en SR de SN en que, aunque en ambos suelos hubo genes relacionados al desarrollo de los frutos, como se comentó anteriormente, en SN la importancia incluyó la presencia y regulación de genes de la fotosíntesis y la clorofila que influye en la pigmentación y en la biosíntesis de azúcares y metabolitos secundarios; por otro lado, en SR se observa una expresión relacionada con genes desde la etapa de floración, y en la maduración de frutos para etapas adelantadas, incluyendo el metabolismo de la pared celular, y el ablandamiento y caída de frutos maduros, lo que podría indicar que SR es favorable para la cosecha y para tener frutos más suaves. Esto podría haber influenciado en que en SN hay mayor proporción de frutos de 45 DPA en etapa de cambio de coloración, y en SR se hayan obtenido más frutos de mayor tamaño, aun cuando la producción total no se vea afectada.

El análisis de enriquecimiento mostró que la mayoría de las rutas estuvieron favorecidas o sobrerrepresentadas por el suelo negro. Entre éstas, funciones metabólicas esenciales, como la fotosíntesis, donde se encuentran proteínas antena de unión a clorofila y otros pigmentos receptores de luz de los LHC y el metabolismo de la porfirina, que lleva a la clorofila; la ruta de las pentosas fosfato (la generación NADPH y pentosas para la biosíntesis); el proteasoma; el ciclo circadiano; metabolismo del nitrógeno y de aminoácidos, fueron afectadas por el tipo de suelo. Los resultados del análisis de enriquecimiento coincidieron con los resultados de genes muy expresados relacionados con la fotosíntesis y presentes en SN mencionados anteriormente. Todo esto apoya que el suelo negro es más favorable para el desarrollo

metabólico y fisiológico de los frutos. Asimismo, algunos otros genes también estuvieron favorecidos en SN, particularmente en el metabolismo y biosíntesis de metabolitos secundarios de interés (terpenoides, carotenoides, y aminoácidos). Si se toma en cuenta que muchos de éstos genes tienen propiedades que determinan la calidad de los frutos, para beneficiar la salud de los consumidores o en características físicas que pueden hacer atractivo al fruto como el color (carotenoides) (Antonio *et al.*, 2018). Los resultados de este trabajo sugieren que el tipo de suelo en el que crecen las plantas es importante para tener frutos con un más rápido y mejor desarrollo, importancia que ha sido resaltada por otros autores (Das *et al.*, 2016). Respecto a la ruta de biosíntesis de capsaicinoides, indirectamente algunas de las vías enriquecidas están relacionadas, como es el caso del metabolismo de glutamato (Zhang *et al.*, 2020), el del metabolismo del ácido 2-oxocarboxílico, donde se puede relacionar con el metabolismo de los aminoácidos de cadena ramificada precursores valina y leucina (Thiele *et al.*, 2008), o la síntesis de antocianinas, con los que comparten intermediarios en la ruta de los fenilpropanoides (Liu *et al.*, 2018). En general, estas vías metabólicas activadas parecen estar más relacionadas con el desarrollo de los frutos que con la acumulación relativa de capsaicinoides, particularmente a los 25 DPA.

3.4.4. Análisis de la expresión individual de genes de las rutas metabólicas enriquecidas y de la ruta de síntesis de capsaicinoides

Con el objetivo de ganar conocimiento sobre el papel de los genes individuales sobre el desarrollo del fruto y de la acumulación de capsaicinoides. Se seleccionaron un grupo de genes representativos de las rutas metabólicas enriquecidas y de genes específicamente relacionados con la biosíntesis de capsaicinoides. Los genes de proteínas antena de unión a clorofila (*LHCA2* y *LHCB3*) de los complejos LHC estuvieron altamente expresados en SN (ambos en el clado I). Estas proteínas, además de tener un papel importante en la captación de la energía de la luz en los LHC y su transferencia hacia los centros de reacción en la fotosíntesis en los fotosistemas I y II (PSI y PSII), también participan en la fotoprotección (Lokstein *et al.*, 2021) respuesta de las plantas a estímulos y estrés ambientales (Zhao *et al.*, 2020; Silva *et al.*, 2016). Su regulación junto con otros genes (Borovsky *et al.*, 2019) y enzimas (Rödiger *et al.*, 2021) de la fotosíntesis y metabolismo de clorofila y de carotenoides también es importante durante la maduración del fruto y para que se den los cambios de coloración. Los ortólogos de *DXS* y *DXS2* (biosíntesis de terpenoides) fueron más expresados en SN, aunque en distintos niveles, lo que sugiere una regulación diferente. Esto concuerda con lo reportado en la literatura para diferentes ortólogos de *DXS*, donde, aunque ambas *DXS* está relacionadas con el metabolismo

de carotenoides, *DXS1* está también relacionada con el desarrollo en las plantas (Zhang *et al.*, 2020b; García-Alcázar *et al.*, 2017), mientras que *DXS2* con el metabolismo secundario, incluyendo a los carotenoides (Zhang *et al.*, 2020b). Las *DXS* están al inicio de la ruta de síntesis de terpenoides, que precede la ruta de síntesis de carotenoides. Tomando esto en cuenta, la diferencia de expresión de los ortólogos de *DXS* podría indicar una posible diferencia en la síntesis de carotenoides, aunque a 25 DPA aún no se observen diferencias en *PDS* y *PSY2*, pero SN tuvo enriquecimiento favorecido en las rutas de terpenoides y carotenoides, y la última adquiere más importancia en etapas más avanzadas de desarrollo de *Capsicum* (Rödiger *et al.*, 2021; Martínez-López *et al.*, 2014), por lo que fue esperada su baja expresión a 25 DPA.

Dentro del clado I también estuvo altamente expresado un ortólogo de *KARI*, una enzima clave en la síntesis de aminoácidos de cadena ramificada (Leung y Guddat, 2009). Su alta expresión en SN junto a genes de proteínas de la fotosíntesis sugiere una regulación relacionada, donde la fotosíntesis se aprovecha para la síntesis de valina y leucina (Chen *et al.*, 2018); además que estos aminoácidos son precursores de los ácidos grasos y de los capsaicinoides. Esto sugiere que en los frutos de 25 DPA en SN la síntesis de valina y leucina podría favorecerse más que en SR, sabiéndose que los tejidos placentarios de chile habanero tienen la capacidad de producir estos aminoácidos (Baas-Espinola *et al.*, 2016). También estuvieron más expresados en SN ortólogos de *GSA* y *GluTR*, ambos de enzimas importantes de la síntesis de porfirinas, hemo y clorofilas, y para que se dé la fotosíntesis (Agrawal *et al.*, 2020; Song *et al.*, 2016), y ortólogos de la *G6PDH* y *6PGD* en la ruta oxidativa de las pentosas fosfato (Nelson y Cox, 2014).

El comportamiento de los genes diferencialmente expresados, concuerda con el resultado de enriquecimiento de las rutas metabólicas relacionadas con el ácido-2-oxocarboxílico, síntesis de terpenoides, la fotosíntesis y la de las pentosas fosfato en SN, y como se ha mencionado anteriormente, sugiere que en el suelo negro se estimula el proceso de fotosíntesis y la biosíntesis, al menos en los frutos de estado de desarrollo intermedio, que llevaría al desarrollo de frutos con mayor contenido de metabolitos como los azúcares, carotenoides y aminoácidos, implicando obtener frutos con mejor valor nutritivo, que es algo que se ha observado en otros estudios en suelos con alto contenido de materia orgánica, nitrógeno y fósforo y diferentes propiedades físicas (Das *et al.*, 2016), aunque esto también depende del tipo de fertilización (Das *et al.*, 2016; Medina-Lara *et al.*, 2008) y el aprovechamiento real de los nutrientes del suelo (Medina-Lara *et al.*, 2019).

Los genes de la ruta de biosíntesis de capsaicinoides no tuvieron diferencia de expresión entre tipo de suelo. Se agruparon en su mayoría en el clado 4 (*AT3-D1*, *KAS*, *KR* y *FAT*), lo que sugiere una regulación conjunta, mientras que *pAMT* en el clado 5, y fue el que tuvo mayor expresión, sugiriendo una regulación en un nivel diferente de los otros genes de la ruta. *AT3-D1* y *KR*, por su parte, fueron el segundo y tercer genes con expresión más alta dentro de su grupo, teniendo una separación de *KAS* y *FAT*. *AT3-D1* y *pAMT* son considerados los genes más específicos de la ruta, ya que se dirigen directamente hacia la síntesis de capsaicinoides (Arce-Rodríguez y Ochoa-Alejo, 2019), y los antecedentes muestran que su expresión suele ser alta al ser muy activos en etapas intermedias de desarrollo de frutos. (Tanaka *et al.*, 2019; Stellari *et al.*, 2010). Por su parte, la importancia de la expresión de *KR* en la determinación del contenido de capsaicinoides y los grupos acilo precursores se ha reportado previamente en *C. annuum* (Koeda *et al.*, 2019), por lo que su alta expresión fue algo esperado. La separación entre *pAMT* y los demás genes puede deberse al hecho de que *pAMT* sintetiza la vanillilamina (Weber *et al.*, 2014) del lado de la ruta de los fenilpropanoides (Arce-Rodríguez y Ochoa-Alejo, 2019), mientras que *KR*, *KAS* y *FAT* participan en otra parte de la ruta, la formación de los grupos acilo que forman a los capsaicinoides (Arce-Rodríguez y Ochoa-Alejo, 2019; Koeda *et al.*, 2019), y *AT3* cataliza la condensación final de ambos sustratos (Ogawa *et al.*, 2015).

El gen *AT3-D1* además es importante en la regulación de la ruta de síntesis de capsaicinoides, incluyendo la expresión de otros genes de la vía, como *KAS* y *pAMT* (Arce-Rodríguez y Ochoa-Alejo, 2015), lo que sugiere una relación entre la expresión de *AT3-D1* y *pAMT*, pese a que haya en clados separados en el presente trabajo. Algunos antecedentes muestran que la expresión de *AT3* y *KAS* es alta solo en variedades picantes de Chile, pero en *pAMT* fue similar en Chile picantes y no picantes (Kim *et al.*, 2014). También se ha demostrado que en variedades no picantes de *Capsicum* y de *C. chinense* con alelos no funcionales de *pAMT*, la expresión de *AT3* y otros genes de la ruta de síntesis de capsaicinoides fueron altas, y los frutos produjeron capsaicinoides (Tanaka *et al.*, 2018; Tanaka *et al.*, 2015), análogos no pungentes de los capsaicinoides (Park *et al.*, 2015a; Tanaka *et al.*, 2010). La información anterior nos indica que, si bien la expresión de *AT3-D1* y *pAMT* tienen una relación en su regulación hasta cierto punto independiente, el hecho de que pertenezcan a la ruta en pasos clave consecutivos indica un grado de relación, y la variedad de Chile junto a las condiciones a las que se someta la planta pueden llegar a hacer que se influencie la expresión entre ambas, por ejemplo, modificando la expresión de *AT3*. La regulación de los ortólogos de *KR*, *KAS* y *FAT* también es importantes dentro de la ruta de capsaicinoides por su participación en la

síntesis de los grupos acilo, ya que existe evidencia de un control de la proporción de los grupos acil-CoA y acil-ACP y la proporción de capsaicinoides correspondientes (Thiele *et al.*, 2008), conduciendo a que en la naturaleza la capsaicina y la dihidrocapsaicina sean los más abundantes.

Dentro del clado 4 se encontraron además ortólogos de *G6PDH* y *RPT4*, que fueron además más expresados en SN. Se ha encontrado que existe una posible relación entre G6PDH, el NADPH y la actividad antioxidante de los frutos que pudiera proteger a los capsaicinoides del catabolismo por POD (Palma *et al.*, 2020). La importancia de esto es que puede sugerir que a los 25 DPA el mayor contenido de capsaicinoides en SN se deba en parte a un menor catabolismo de capsaicinoides que en SR, favoreciendo así la mayor acumulación de éstos en los suelos negros (Zamudio-Moreno *et al.*, 2014). Además, parte importante de la actividad de ácido graso sintasa que participa en la ruta de síntesis de capsaicinoides implica una reacción de reducción catalizada por la KR con participación del NADPH (Koeda *et al.*, 2019), que pudiera relacionarse con la actividad de la G6PDH del ortólogo hallado en el presente trabajo. *RPT4* por su parte es un ortólogo de la subunidad S10B del proteasoma 26S. Su relación con los capsaicinoides pudiera estar relacionado con los cambios en las paredes y membranas de las células y con el transporte vesicular y/o posterior almacenamiento en vacuolas en las células donde se sintetizan los capsaicinoides (Tyler *et al.*, 2016; Zamski *et al.*, 1987). Esto muestra que no solo es importante el estudio de los genes de la síntesis de capsaicinoides, sino también de los procesos biológicos relacionados, como el transporte y almacenamiento de estos metabolitos, y los cambios celulares relacionados.

En clado 3 se encontró a *AT3-D2*, que otros autores han descrito como un parálogo de *AT3-D1*, de los cuales se conoce que cada una tiene distinto nivel de expresión en frutos de *Capsicum*, siendo *AT3-D1* el más expresado en las variedades picantes (Qin *et al.*, 2014), lo que concuerda con lo observado en el presente trabajo. *MYB*, por su parte, al igual que los genes de la ruta de síntesis de capsaicinoides, no tuvo diferencia de expresión entre suelos. Éste fue un ortólogo de factores de transcripción MYB, cuya regulación es importante en el control de la expresión de genes de la ruta de síntesis de capsaicinoides, como *KAS*, *FAT*, *pAMT* y *AT3*, así como el nivel de capsaicinoides de las variedades picantes de Chile (Zhu *et al.*, 2019). Se ha reportado que los genes *pAMT*, *KAS* y *AT3* tienen sitios de unión para factores MYB (Sun *et al.*, 2020, Sun *et al.*, 2019), que regula la expresión de genes en respuesta a factores ambientales (Arce-Rodríguez y Ochoa-Alejo, 2017), por lo que el control de la expresión de estos factores de transcripción entre tipos de suelo sería interesante para tratar de modificar el contenido de

capsaicinoides. La expresión diferencial de los genes *AT3*, su relación con otros genes de la síntesis de capsaicinoides como *KAS* y *pAMT*, y finalmente su regulación con la síntesis de capsaicinoides en las plantas de *C. chinense* cultivadas en diferentes tipos de suelo deber ser más investigada.

Otros ortólogos analizados que pudieran relacionarse con la síntesis de capsaicinoides fueron *KARI*, *GS* y *GOGAT*, donde se ha observado por su relación con el metabolismo del nitrógeno y en la síntesis de valina y leucina. Además, para *GS* y *GOGAT*, por su importancia en el metabolismo del nitrógeno, tienen relación con la expresión de genes y actividad de enzimas de la ruta de síntesis de capsaicinoides y con el contenido de capsaicinoides en frutos de chile (Zhang *et al.*, 2020a). Aun cuando no haya habido diferencia en la expresión de *GS* y *GOGAT*, conociendo la importancia del contenido de nitrógeno en las diferencias que existen entre los suelos negros y los suelos rojos y en el metabolismo de capsaicinoides, la regulación de los genes como *GS*, *GOGAT* y *KARI* también sería interesante como posibilidad para modificar el contenido de capsaicinoides entre suelos.

3.5. CONCLUSIONES

Las plantas de chile habanero crecidas en SN y SR presentaron algunas diferencias en el fenotipo de sus frutos: mayor número de frutos de 45 DPA en estado de madurez intermedio (cambio de coloración) y menor número de frutos inmaduros (verdes) en SN; y mayor número de frutos más grandes en SR. A nivel transcriptómico, SN tuvo enriquecidas muchas rutas relacionadas con la fotosíntesis, la ruta de las pentosas fosfato, el proteasoma, el metabolismo de aminoácidos y metabolitos secundarios, incluyendo los capsaicinoides. Esto puede indicar un favorecimiento de la calidad del fruto en los suelos negros relacionado con su contenido nutritivo de metabolitos que favorezcan la salud, destacando el favorecimiento de la expresión de genes relacionados con la fotosíntesis y la clorofila. Los genes de la ruta de síntesis de capsaicinoides fueron altamente expresados a los 25 DPA, pese a no tener diferencia en expresión entre suelos. Sin embargo, SN alcanzó un mayor contenido de capsaicinoides a los 25 DPA que mantuvo hasta los 45 DPA, mostrando ser favorable para una obtención más temprana de un máximo contenido de capsaicinoides que podrían usarse para la extracción y uso en la industria. En SN, rutas y actividades relacionadas con el metabolismo de capsaicinoides tuvieron genes más expresados, como la síntesis de valina y leucina, el proteasoma (almacenamiento vesicular de capsaicinoides) y la ruta de las pentosas fosfato que podría tener un efecto antioxidante hacia los POD que degradan los capsaicinoides. Dentro de

los genes de la ruta de síntesis de capsaicinoides, se encontraron dos homólogos del gen de la CS (*AT3-D1* y *AT3-D2*) que demostraron tener diferente nivel de expresión entre sí, lo que sugiere una regulación diferente para la síntesis de capsaicinoides. En conjunto, nuestros resultados muestran evidencia de efectos transcriptómicos importantes que influyen en el desarrollo y metabolismo de frutos de chile habanero, y abre la puerta a profundizar los efectos de estas diferencias en el fenotipo y calidad de los frutos.

CAPÍTULO IV

Este capítulo corresponde al artículo sometido para publicación: **Haplotype analysis, regulatory elements and docking simulation of structural models of different AT3 copies in the genus *Capsicum*** en la revista **Tropical Plant Biology** con fecha de sometimiento del 20 de junio de 2022 y con los autores Burgos-Valencia, E., Narváez-Zapata, N., Echevarría-Machado, I., Ortega-Lule, G., Medina-Lara, F., Martínez-Estévez, M.

HAPLOTYPE ANALYSIS, REGULATORY ELEMENTS AND DOCKING SIMULATION OF STRUCTURAL MODELS OF DIFFERENT AT3 COPIES IN THE GENUS *Capsicum*

ABSTRACT

Capsaicinoids are responsible of the pungency in *Capsicum* species. These are synthesized by the capsaicinoid synthase (CS) encoded by the *AT3* gene, which catalyzes the transference of an acyl moiety from a branched-chain fatty acid-CoA ester to the vanillylamine to produce capsaicinoids. Some *AT3* gene copies have been identified on *Capsicum* genome. The absence of capsaicinoids in some non-pungent accessions are relate to mutant *AT3* alleles. The polymorphism between CS protein copies can affect the tridimensional structure of the protein and the affinity for its substrates, and this could affect the fruit pungency. In this study, 32 *AT3* sequences covering *Capsicum* pungent and non-pungent accessions were characterized. These were clustered in AT3-D1 and AT3-D2 groups and representative sequences were analyzed. Genomic upstream analysis shows different regulatory elements, mainly responsive to light and to abiotic stress. Amino acid substitutions close to the predictable HXXXD and DFGWG motifs were also identified. *AT3* sequences were modeled showing a BAHD acyltransferase structure with two connected domains. A pocket with different shape, size and composition between *AT3* models was found inside the protein, with the conserved motif HXXXD exposed to it, and a channel for their accessibility. CS substrates exhibits high interaction energies with the His and Asp conserved residues. *AT3* models have different interaction affinities with the (E)-8-methylnon-6-enoyl-CoA, 8-methylnonanoyl-CoA and vanillylamine substrates. These results allow suggested that AT3-D1 and AT3-D2 sequences encode CS enzymes with different regulatory factors and substratum affinities.

Keywords: Capsaicinoid synthase, *Capsicum*, Regulatory factors, Protein modeling

4.1. INTRODUCTION

Capsicum genus covers 30 species including domestic species as *C. annuum*, *C. baccatum*, *C. frutescens*, *C. chinense* and *C. pubescens*. They are also known as peppers (Barboza *et al.*, 2019; Carrizo García *et al.*, 2016; Bosland and Votava, 2012). These are appreciated by their sensorial properties as taste, smell and color that make them attractive and useful in gastronomy. Also, these are a source of multiple nutrients and beneficial metabolites, mainly capsaicinoids (Ananthan *et al.*, 2018; Antonio *et al.*, 2018; Aranha *et al.*, 2017). Capsaicinoids are a group of compounds responsible of the pungency of the pepper fruits, which have diverse medical (analgesic, anti-inflammatory effect, antioxidant, anticancer, antitumoral and gastrointestinal benefices) and industrial (sauces, condiments, aerosols, chemical defense devices, cosmetics, repellents and antimicrobials) applications (Ghiasi *et al.*, 2019; Lin *et al.*, 2018; Clark and Lee, 2016; Srinivasan, 2016; Aza-González *et al.*, 2011; Caterina *et al.*, 1997). Generally, the capsaicin and the dihydrocapsaicin are the mayor components of the capsaicinoid composition in pepper fruits, although the proportions vary depending on the developmental stage of the fruit and the pepper variety (Olguín-Rojas *et al.*, 2019; Benmansour-Gueddes, 2015). The capsaicinoids are synthesized and accumulated mainly in the epidermal cells of the interocular septum of the fruit (Tyler *et al.*, 2016; Stewart Jr *et al.*, 2007; Zamski *et al.*, 1987). The chemical structure of capsaicinoids is composed of a vanillyl group bonded by and amide group to an alkyl chain. The different types of capsaicinoids differ by the alkyl chain in length, saturations and the presence of branched or linear chains (Antonio *et al.*, 2018). The capsaicinoids are synthesized by the conjunction of the phenylpropanoid pathway that gives vanillylamine, and the metabolism of branched-chain fatty-acids that gives the acyl moiety. In the current pathway model, the final step of the pathway is a condensation of the vanilloid and acyl moieties that gives the capsaicinoid (Arce-Rodríguez and Ochoa-Alejo, 2019). It has been proposed that this reaction is catalyzed by the enzyme CS, which is an acyltransferase encoded by the *AT3* gene (*Pun1* locus) localized in the chromosome 2 (Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005). *AT3* gene codified a putative acyltransferase of BAHD (Benzylalcohol O-acetyltransferase, anthocyanin O-hydroxycinnamoyltransferase, anthranilate-N-hydroxycinnamoyl/benzoyltransferase, deacetylvindoline 4-O-acetyltransferase) superfamily (Stewart Jr *et al.*, 2005) with expression in *Capsicum* fruits, closely following the capsaicinoid accumulation in medium development stages and green mature fruits (Martínez-López *et al.*, 2014; Qin *et al.*, 2014). The absence of capsaicinoids in some non-spicy varieties are relate to

mutant *AT3* alleles (Kirii *et al.*, 2017; Stellari *et al.*, 2010; Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005).

Despite the previous information about *AT3* gene and its contribution in the capsaicinoid synthesis in fruits, there is little information about of their encoded CS protein, its activity or its substrate affinity. Also, it is known that *AT3* gene have different copies in the *Capsicum* genome (Egan *et al.*, 2019; Kim *et al.*, 2014)) but few is known on the differences between *AT3* gene copies. Most of the studies are focused in one copy or in the *AT3* gene without considers the difference between the genomic copies. Although, it has been observed that they are differentially expressed between varieties and within the same variety, and its regulation is different between pungent and non-pungent cultivars (Qin *et al.*, 2014).

The CS activity together with *AT3* gene expression has been correlated with the capsaicinoid accumulation in fruits (Zhang *et al.*, 2020a). It has been observed that enzymatic extracts of placental tissues in *C. annuum* fruits have CS activity using as substrates vanillylamine and acil-CoA derived from C9 and C10 saturated and unsaturated fatty acids and producing nordihydrocapsaicin, capsaicin, dihydrocapsaicin and homodihydrocapsaicin (Fujiwake *et al.*, 1980). Also, there is indirect evidence with the use of antibodies against the CS protein indicating that *AT3* down-regulation is related with the reduction of the corresponded protein and of the capsaicinoids in placental tissues (Ogawa *et al.*, 2015). CS protein affinities can be predicted by molecular docking of the enzyme with different substrates as acyl-CoA and vanillylamine that form the capsaicinoids (Ferreira *et al.*, 2015). The polymorphism between CS protein copies can affect the tridimensional structure of the protein and the affinity for its substrates, and this could affect the fruit pungency (González-Zamora *et al.*, 2020). Therefore, current study was made to explore the contribution of different *AT3* copies identified on the *Capsicum* genome on the capsaicinoids synthesis by comparison of their different regulatory regions and by models their tridimensional structures in order to make a docking comparison with their possible substrates.

4.2. MATERIAL AND METHODS

4.2.1. *AT3* orthologues identification and characterization

AT3 and acyltransferase sequences of *C. annuum* (AY819029.1) and *C. chinense* (AY819027.1) were used in the TBLASTX and BLASTN tools (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) on the Transcriptome Shotgun Assembly sequences,

available in the NCBI, and on the genomes of *C. chinense*, *C. annuum* and *C. baccatum* plants to search gene orthologues within these species. Also, a BLAST search was made in the NCBI database to find additional sequences in the *Capsicum* genus. Sequences with high e-values ($<1e-100$), and that includes the HXXXD, DFGWG and BAHD acyltransferases conserved motifs (Stewart Jr. *et al.*, 2005) were selected. These nucleotide sequences were aligned by ClustalW and a phylogenetic tree was constructed by Maximum likelihood with a bootstrap of 1000, using the software Mega v.7.0.26 (Kumar *et al.*, 2016). In addition, *C. annuum* (AAV66311.1) and *C. chinense* (AAV66309.1) protein sequences were blasted with the non-redundant (BLASTP) protein database, to find *Capsicum* gene orthologues with the same characteristics previously described. Respective aminoacidic sequences when these were available, or the predictable open reading framers obtained by using the Open Reading Frame Finder tool (<https://www.ncbi.nlm.nih.gov/orffinder/>), were aligned by ClustalW together with other protein sequences of closely related species as *Solanum pennellii*, *S. tuberosum*, *S. lycopersicum*, *Datura stramonium*, *Petunia axillaris*, *Lycianthes dejecta* and *Nicotiana tabacum*. A phylogenetic tree was constructed by Maximum likelihood with a bootstrap of 1000, using the software Mega v.7.0.26. Trimmed on some protein sequences was conducted to adjust the sequences with the protein range 36 aa to 415 aa for AT3-D1 group, for cba02 (PHT26523.1) the range 36 aa to 417 aa, and the protein range 36 aa to 416 aa for the others AT3-D2 group sequences and were aligned by ClustalW using Mega v.7.0.26. A haplotype map was made with the software Network v.10.2.0.0 (<https://www.fluxus-engineering.com/>) with the algorithms Reduced Median-Median Joining to identify protein groups. The multiple alignment representations were made using the software CLC Sequence Viewer v.8.0 (<https://digitalinsights.qiagen.com/>).

4.2.2. Genomic regulatory elements

A search of regulatory and promoter elements was made in the representative sequences of *C. annuum* Zunla-1 can03 and can06 and *C. chinense* PI159236 cch02 and cch03 of the main haplotype groups. A region 2000 pb upstream, on the chromosome 2 of the *C. annuum* Zunla-1 (CM002813.1) and *C. chinense* PI159236 (CM008432.1) genomes, of the start codon (ATG) of these sequences was selected. Regulatory elements search was made in the platform PlantCARE (Lescot *et al.*, 2002) in each sequence, and the elements were compared by stimuli and presence/absence of the element between haplotype regulatory regions.

4.2.3. Modeling the AT3 protein

The I-TASSER platform was used to obtain the models of representative sequences of the different groups founded in the haplotype analysis (Yang *et al.*, 2015). This platform was selected since allows modeling structures by a combination of modeling by threading and ab initio methods, searching in the Protein Data Bank remote homologues and using them as templates (Zheng *et al.*, 2019; Roy *et al.*, 2010). The best model generated was evaluated based in the C-Score, the T-score and the Root-Mean-Square Deviation RMSD by the platform. The Ramachandran diagram was analyzed to search residues with inadequate torsions. Additionally, an alignment of the models with the templates and homologue proteins was made, and the motifs, secondary structures and tertiary structures were compared. The tridimensional representations were generated with the PyMOL v.2.3.0 (The PyMOL Molecular Graphics System, Version 2.3.0 Schrödinger, LLC) and Swiss PDB Viewer Program v.4.1.0 (Guex and Peitsch, 1997), and the alignments between structures and Ramachandran plots were generated with the Swiss PDB Viewer Program v.4.1.0. The pockets of the AT3 proteins were determined with the platform of Protein Plus (Schöning-Stierand *et al.*, 2020) with the DoGSiteScorer (Volkamer *et al.*, 2012). The structures of the AT3 protein were refined using the VMD software v1.9.3 with the extension QwikMD (Humphrey *et al.*, 1996). The followed protocol was a minimization of 2000 steps at 40 °C, an annealing stage of 14000 steps at 35° C, an equilibration stage of 627500 steps at 30 °C and finally a molecular dynamic of 5000000 steps at 25 °C without restrictions in this last stage. Each step had a time of 2 fs. The protocol was realized with a NVT ensemble with an implicit solvent and a pressure of 1 atm in with the default parameters of the program.

4.2.4. Molecular docking of the AT3 protein

An analysis of docking was made with the AT3 models selected and the proposed substrates of the capsaicin synthase in the model of the biosynthesis of capsaicinoids. The 2D structure of the molecules used as ligands were downloaded from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and the CheBI database (<https://www.ebi.ac.uk/chebi/init.do>). The ligands 8-methylnonanoyl-CoA (PubChem CID 154586165) and (E)-8-methylnon-6-enoyl-CoA (PubChem CID 146026638) were used as they are the respective substrates for the formation of dihydrocapsaicin and capsaicin, the most abundant capsaicinoids in chili fruits (Fujiwake *et al.*, 1980). The capsaicinoids biosynthesis substrate vanillylamine (PubChem CID 70966) was also assayed. Also, due to the close

relationship with the CS, two specific deacetylvindoline 4-O-acetyltransferase (DAT) substrates were also assayed: the 4-O-deacetylvindoline (PubChem CID 129320270) and the acetylCoA (CheBI ID 57288).

The ligand structures were minimized with the software Avogadro v.1.2.0 using the force field mmf94 (Hanwell *et al.*, 2012). The docking was made with the software Autodock Vina v1.1.2 with AutodockTools v1.5.6 (Trott and Olson, 2010). First, the AT3 structures were prepared by aggregating the polar hydrogens and the Kollman charges were calculated. The non-polar hydrogens were eliminated (merged). The ligands were used after the previous minimization in Avogadro. A rigid docking was made for each ligand with a Grid Box in the region of the solvent channel of the acyltransferase that include inside the motif HXXXD of the acyltransferases. The exhaustiveness was 8 in all the dockings and the energy range of the binding modes was 3. The best conformation (the lower energy mode) was analyzed. The representations of the conformations were made with the PyMOL v.2.3.0 software.

4.3. RESULTS

4.3.1. AT3 gene analysis

Sixty-six closest sequences to *AT3* were identified within the *Capsicum* genus (Supplementary Table 4.1). *AT3* analysis included acyltransferases of closely related species. Sequence alignment allows identify 2 motifs conserved among the acyltransferases of the superfamily BAHD. One is the motif HXXXD, where X is any aminoacid and that occurs as HKIGD in the *AT3* acyltransferases, and the other is the conserved motive DFGWG (Supplementary Figure 4.1). *AT3* gene has one intron and two exons. From the sixty-six initial sequences, the majority of the sequences belonged to the *C. annuum* specie and included spicy (Thai Hot, Hot1493, CM334, Tabasco, Habanero and Zunla-I) and non-spicy (Jupiter, Maor, Sweet3575, ECW and NMCA30036) cultivars (Supplementary Table 4.1). In total, 35 *Capsicum* sequences, including those of both, pungent and non-pungent cultivars, were selected to protein phylogenetic analysis, discarded all sequences with stop codons, CDS and pseudogenes. In addition, acyltransferases of other Solanaceae species as *D. stramonium*, *P. axillaris* subsp. *parodii*, *S. lycopersicum*, *S. pennelli*, *S. melongena* and *L. dejecta* were included in this analysis with comparison aim. *Capsicum* *AT3* sequences were grouped in a clade called “AT3-like” divided in two subclades (AT3-D1 and AT3-D2; Figure 4.1).

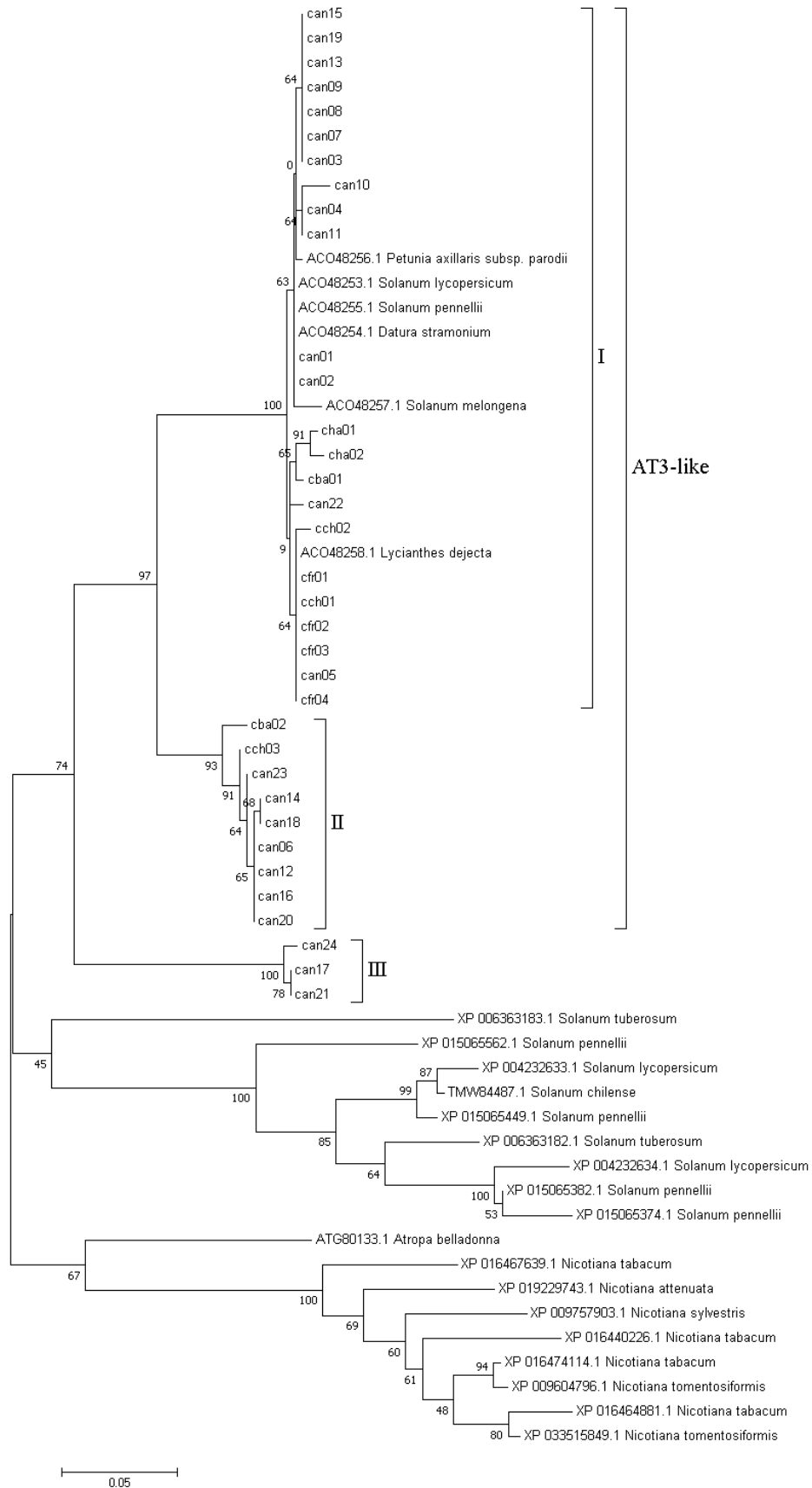


Figure 4.1. Phylogenetic tree of the AT3 sequences in *Capsicum* and homologues. The sequences relatives to the AT3 protein were searched by BLAST in the NCBI database. A multiple alignment was made by ClustalW and the tree was constructed by Maximum likelihood with a bootstrap of 1000 using the software Mega v.7.0.26.

Most of the non-*Capsicum* sequences were clustered distant from the AT3-like clade. In the AT3-D1 subclade, most *C. annuum* (can) sequences were separated from *C. chacoense* (cha), *C. baccatum* (cba), *C. frutescens* (cfr) and *C. chinense* (cch) sequences. Only can22 and can05 (*C. annuum*) was grouped together to other *Capsicum* species. Most sequences of the AT3-D1 subclade have 1671 pb gene length (including introns), except cba01 that has 1674 pb, and have a CDS of 1323 pb and 440 of translate aa (Supplementary Table 4.1). The AT3-D1 sequences of the phylogenetic analysis of this clade are formed mostly by pungent cultivars. Subclade AT3-D2 was formed by only 9 sequences, being the most AT3 acyltransferase proteins. Although, there are some sequences (can14 and can16) annotated as hypothetical DAT. Most the sequences in the subclade AT3-D2 have a CDS of 1326 pb and 441 of translate aa, except cba02, which had a CDS of 1329 and 442 aa of protein sequence. Genomic sequences in AT3-D2 subclade were bigger than those of the AT3-D1 subclade due to their longer intron sequence. AT3-D2 clade has sequences of some non-pungent cultivars as ECW (can14). In general, all the sequences in these two subclades code a putative BAHD acyltransferase domain and therefore are considered as acyltransferase proteins (Supplementary Figure 4.1). Finally, a third subclade was identified. This is formed by only 3 sequences: can17, can21 and can24, with can17 annotated as a dehydrogenase/decarboxylase-like protein (KAF3637600.1). These sequences shared the motifs HXXXD and DFGWG but are closely related with the 3- β -hydroxysteroid-dehydrogenase/descarboxylase, which were also discarded to further analysis (Supplementary Table 4.1, Figure 4.1).

4.3.2. Haplotype network analysis

Translated sequence analysis was conducted with 32 *Capsicum* AT3 sequences. Haplotype network analysis divided the sequences into two main groups (Figure 4.2). Group AT3-D1 was the most abundant and polymorphic, formed by 23 sequences including five species: *C. annuum*, *C. baccatum*, *C. frutescens*, *C. chinense* and *C. chacoense*. The accessions *C. annuum* can04 (QCT85208.1) and can11 (QCT85209.1), previously annotated as AT3-D1 (Qin *et al.*, 2014), belonged to this group. In general, most of the sequences within this group were

subgrouped according to their specie, although, the can05 (*C. annuum*) sequence was subgrouped with some *C. frutescens* (cfr01-cfr03). This sequence has an identity >97.8 % with regard to other *C. annuum* sequences. In general, the identity among accessions of AT3-D1 group was >96%. The group AT3-D2 is formed by nine sequences of the *C. annuum*, *C. chinense* and *C. baccatum* species. Can12 sequence annotated as AT3-D2 (Qin *et al.*, 2014) belonged to this group. Only the sequence of *C. baccatum* (cba02) shows a high separation (11 polymorphic sites) and an identity between 95.8-96.8% with the others AT3-D2 sequences. The identity among the AT3-D2 sequences was >95%. In general, between AT3-D1 and AT3-D2 groups the identity ranged between 85-89% and exhibit at least 36 polymorphic changes (Figure 4.2). Most of the changes occur in the exon 2 of the gene. A conserved GN sequence (Gly302 and Asn303) and the residue Arg289, that are found conserved in many acyltransferases, was also founded in all the AT3 sequences (Supplementary Figure 4.1).

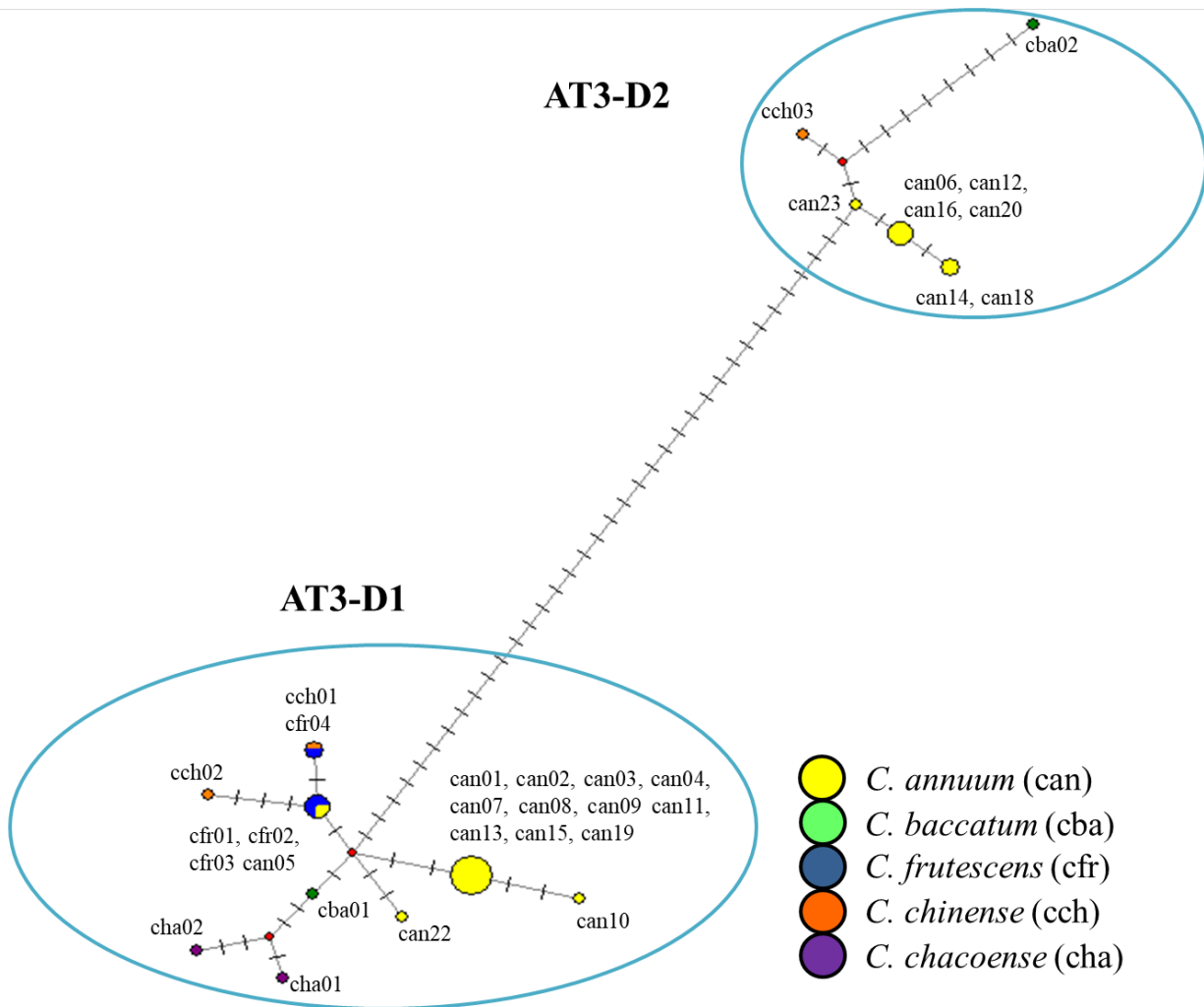


Figure 4.2. Haplotype map of the 32 AT3 sequences in *Capsicum*. The main groups AT3-D1 and AT3-D2 are delimited with a blue circle. The alignment was made with ClustalW with the software Mega v.7.0.26. The haplotype map was made with the algorithm Reduced Median-Median Joining using the software Network v.10.2.0.0.

4.3.3 Regulatory elements analysis

AT3-D1 and *AT3-D2* representative sequences were localized and analyzed on *C. annuum* Zunla-1 (CM002813.1) and *C. chinense* PI159236 (CM008432.1) genomes and the regions 2000 pb upstream from the start codon in the positions CM002813.1:c143429901-143427902 (can03), CM002813.1:c143405095-143403096 (can06), CM008432.1:149591797-149593796 (cch02) and CM008432.1:149616377 to 149618376 (cch03) were analyzed. Regulatory elements analysis was conducted on the upstream region in these genes (Supplementary Table

4.2). Upstream regions in between *AT3-D1* (can03) and *AT3-D2* (can06) was 32.4% of identity and 33.3% of identity between *AT3-D1* (cch02) and *AT3-D2* (cch03). Regulatory elements were organized by stimuli in each sequence (Table 4.1). The stimuli detected were light, drought and dehydration, anaerobiosis, mechanical wounding, pathogenesis, abscisic acid, auxin, gibberellin, methyl jasmonate, ethylene and salicylic acid, heat and salinity. The light was the stimuli with more diverse factors and the abiotic stress the stimuli with higher factor number. The transcriptional factors more abundant (> 5 hits) were ABRE for abscisic acid, Box 4, GT-1 motif and G-box for light, and W-Box, MYB, STRE and MYC for abiotic stress. MBS binding site for drought response was not founded specifically can06 *AT3-D2*, but it was found in cc03 *AT3-D2* and in the *AT3-D1* regulatory region of both species. WRE3 for mechanical wounding was found in *AT3-D2*. Some regulatory elements of response to hormones as abscisic acid (ABRE3a, ABRE4), auxin (AuxRE), gibberellin (GARE-motif, TACT-box), ethylene (ERE), methyl jasmonate (CGTCA-motif), and as-1 for multiple hormones, including salicylic acid, were specific of each *AT3* sequence in at least one specie. F-box, that can be related with floral development, and a circadian cycle motif were found only in *AT3-D1*, and MBSI was found only in cch02. The transcriptional factors with more differential hits in each haplotype were Unnamed1, ABRE for abscisic acid, ARE for anaerobiosis, ERE for ethylene, TCT-motif and G-box for light and STRE for abiotic stress. G-box and STRE were the factors with more differential hits; widely represented in can12 and cch03 (*AT3-D2*), which might suggest that the light and the abiotic stress have an important role in the differential expression between these *AT3* genes.

Table 4.1. Regulatory elements at 2000 pb upstream of the start of *AT3-D1* and *AT3-D2* in *C. annuum* Zunla-1 and *C. chinense* PI159236

Regulatory element	Stimuli	<i>AT3-D1</i> (cch02)	<i>AT3-D2</i> (cch03)	<i>AT3-D1</i> (can03)	<i>AT3-D2</i> (can06)
ABRE	Abscisic acid	2	4	2	4
ABRE4	Abscisic acid	0	1	0	1
ABRE3a	Abscisic acid	0	1	0	1
ARE	Anaerobiosis	2	0	2	0
AuxRE	Auxin	0	0	1	0
TGA-element	Auxin	1	0	0	0

Box4	Light	4	5	4	5
G-box	Light/Multifunctional regulatory element	1	4	1	4
GATA-motif	Light	1	1	1	2
GT1-motif	Light	1	2	1	2
Gap-box	Light	1	0	1	0
Pc-CMACc	Light	0	0	1	0
ACE	Light	0	1	0	1
AE-box	Light	0	1	0	1
GA-motif	Light	0	1	0	1
ATCT-motif	Light	0	0	0	1
MRE	Light	0	1	0	1
TCT-motif	Light	1	1	0	2
GARE-motif	Gibberellin	1	0	1	0
P-box	Gibberellin	1	1	1	1
TATC-box	Gibberellin	1	0	1	0
ERE	Ethylene	2	0	1	0
MBS	Drought	1	1	1	0
CGTCA-motif	MeJA	1	0	1	0
MYC	Drought/MeJA/abiotic stress/defense	3	2	3	2
TGACG-motif	MeJA	1	0	1	0
as-1	Salicylic acid/MeJA/auxin/pathogenesis	1	0	1	0
TCA	Salicylic acid	1	2	0	2
WRE3	Mechanical wound	0	1	0	1
W-box	Abiotic stress/mechanical wound/heat/salinity	2	1	2	2

CAPÍTULO IV

MYB	Abiotic stress	3	3	2	3
STRE	Abiotic stress	1	5	1	4
F-box	Floral development	1	0	1	0
CAT-box	Meristem expression	0	1	0	1
Circadian	Circadian cycle	1	0	1	0
MBSI	Flavonoid biosynthesis	1	0	0	0
AAGAA- motif		2	1	1	1
Unnamed 1		3	0	3	0
Unnamed 4		12	9	11	9
Unnamed 6		0	1	0	0
O2-site	Zein metabolism	0	0	0	1
AT-rich element	Binding site of AT-rich DNA binding protein (ATBP-1)	1	1	1	1
AT-TATA- box	Core promotor element	2	3	4	2
TATA-box	Core promoter element	7	9	7	7
CAAT-box	Common cis-acting element in promoter and enhancer regions	2	2	2	2

4.3.4. Modeling the AT3 protein

Representative can01 (AT3-D1) and can12 (AT3-D2) sequences were selected to carry out the AT3 protein structures. To date, there are not tridimensional structure models of this protein. Therefore, related proteins were selected as templates. The model vinorine synthase of *Rauvolfia serpentina* (PDB ID: 2bgh) and the model hydroxycinnamoyltransferase (HCT) of *Sorghum bicolor* (PDB ID: 4ke4), which are closest to the AT3-D1 and AT3-D2 structures, respectively (Supplementary Table 4.3). The AT3 structures are similar to the general structure of the acyltransferases of the superfamily BAHD, as the vinorine synthase and the HCT (Supplementary Figure 4.2). TM-score and the Root-Mean-Square Deviation (RMSD) of AT3-D1

structure with the vinorine synthase was 0.907 and 0.66 Å. TM-score and RMSD of AT3-D2 structure with the HCT was 0.921 and 1.65 Å (Supplementary Table 4.3). These values suggest a structural similarity of the proteins query with the selected models, particularly for AT3-D1 with the vinorine synthase. AT3-D1 model has secondary structures (α -helices and β strands) highly aligned with the vinorine synthase template. The loops were the regions with lower alignment between both structures, although these regions are difficult to model by homology (Supplementary Figure 4.2). Therefore, this model was selected to further analysis. Here it is important mention that Ramachandran plot of AT3-D1 shows residues in phi and psi angles out of the expected suggesting some class of structure disarrangement. Phi and psi angles were corrected by molecular dynamic generating a refined model. However, the accessibility in the internal cavity was prevented and, therefore, this refined model was discarded to further analysis (data no show).

The model of AT3-D1 shows a C-score of 0.51, a TM-score of 0.78 and a RMSD of 5.9 Å (Table 4.2). AT3-D1 structure has two domains with a similar structure, a core of mixed β -sheets and various helices around them. AT3-D1 has 14 β -strands in a mixture of parallel and antiparallel conformation and 12 helices. The domains are connected with a large loop (Figure 4.3A). Domain 1 is formed by the residues 1 to 204 and residues 378-398, forming eight β -strands and six helices. Six β -strands form a central core in a mixture of parallel and antiparallel conformations and four helices surround the core of β -sheet. Also, there are two helices and two β -strands in an antiparallel conformation in the surface of the protein. This domain includes at conserved motif HXXXD (His169 to Asp173). The conserved motif DFGWG (Asp384 to Gly388) is in a loop that is far away from the conserved motif HXXXD (Figure 4.3A). Domain 2 is formed by residues 227-377 and the residues 399-440. There are six β -strands forming a β -sheet core in a mixture of parallel and antiparallel conformations. Six helices surround the β -strands. The loop that connects both domains is formed by the residues 205-226 (Figure 4.3A). A pocket was found inside the protein, forming a tunnel that pass thought the center of the protein in the region between the two domains where the conserved motif HXXXD is localized. The His169 and Asp173 have the side chains exposed to the internal cavity. The tunnel has two entrances in opposite sides of the protein that would serve for the accessibility of substrates to the central region (Figure 4.3B and 4.3C).

Initial AT3-D2 model show a C-score of 0.15, a TM-score of 0.73 and a RMSD of 6.7 Å (Table 4.2). AT3-D2 structure shows some residues in phi and psi angles out of the expected suggesting some class of structure disarrangement (data no shown). Therefore, a refinement by

CAPÍTULO IV

molecular dynamic was conducted generating a refined model where most of these unexpected phi and psi angles were corrected. Refined structure of AT3-D2 was similar to AT3-D1 model with some differences mainly in the domain 1 which has seven helices and nine β -strands and the conserved His and ASP residues, belonged to conserved motif HXXXD, in the 168 and 172 positions, respectively (Figure 4.3A). Differently to AT3-D1 model, the Asp172 side-chain keeps away from the central region of the cavity (Figure 4.3B and 4.3C). In general, AT3-D1 and AT3-D2 models comparison share 87.04% of identity and in the structural alignment have a RMSD of 1.83 Å. Also, they exhibit differences in the internal cavity, the shape, volume, area and depth of the pocket, with AT3-D1 having a higher volume and area (Figure 4.3 and Supplementary Table 4.4). Also, the pocket shows differences in the aminoacyl composition, having AT3-D2 higher apolar and positive aminoacyl composition, and AT3-D1 having higher polar and negative aminoacyl composition (Supplementary Table 4.4). These differences might affect the affinity for the possible substrates of the capsaicin synthase (acyl-CoA medium chains fatty acids and the vanillylamine).

Table 4.2. General confidence values of AT3-D1 and AT3-D2 models and their docking interaction with different substrates

Confidence values	AT3-D1	AT3-D2
	C-score	0.51
Estimated TM-score	0.78±0.10	0.73±0.11
Estimated RMSD (Å)	5.9±3.7	6.7±4.0

Ligand	Docking energies	
	AT3-D1 (kcal/mol)	AT3-D2 (kcal/mol)
(E)-8-methylnon-6-enoyl-CoA	-8.6	-9.3
8-methylnonanoyl-CoA	-8.6	-8.4
Vanillylamine	-5.1	-5.0
4-O-deacetylvindoline	-8.2	-7.0
Acetyl-CoA	-8.3	-8.9

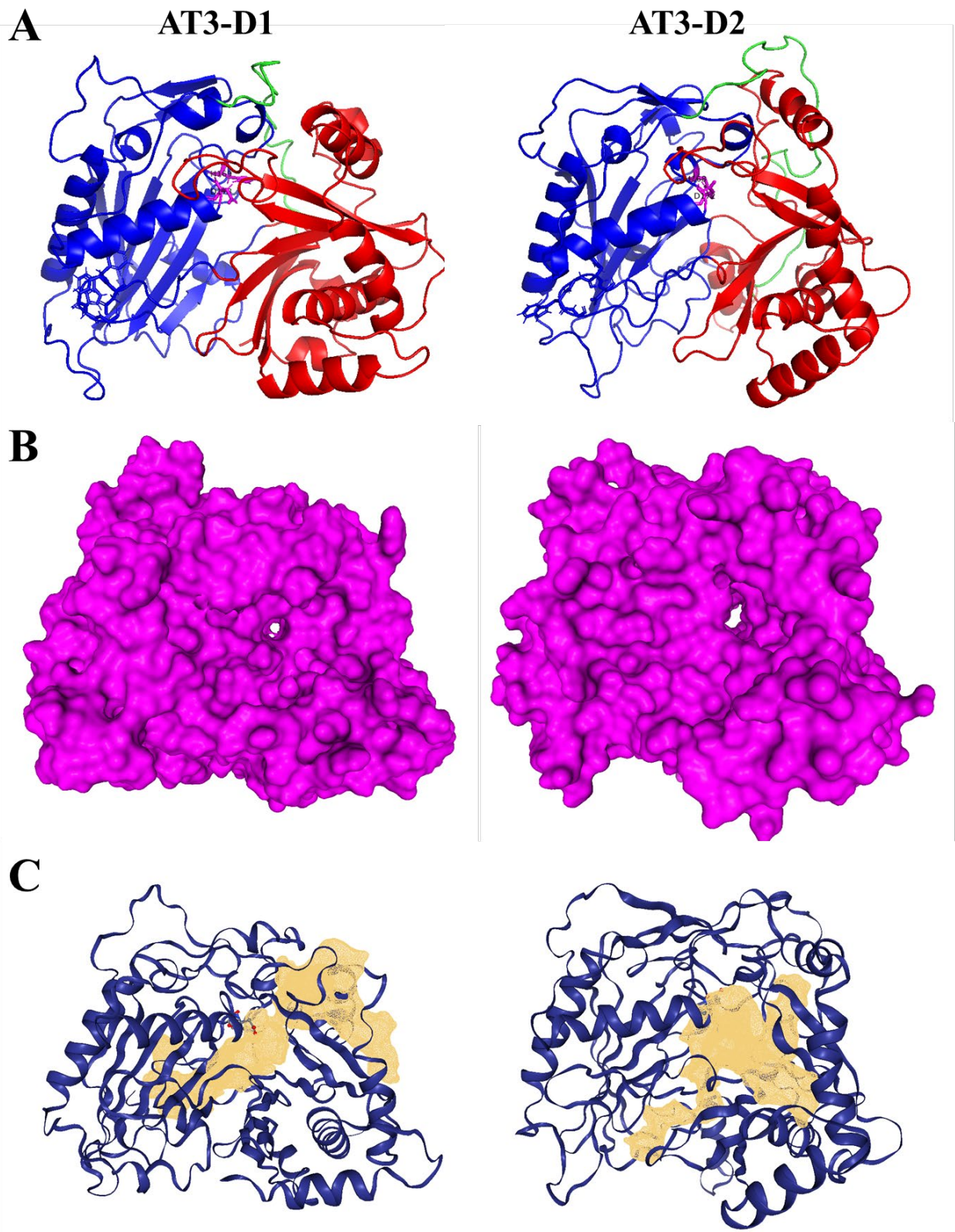


Figure 4.3. Structure models and surfaces of AT3-D1 (left) and AT3-D2 (right). **A** Tridimensional structure of AT3-D1 (left) and AT3-D2 (right). Domains 1 and 2 are in blue and red, respectively. The loop that connects both domains is in green. The conserved H169 and D173, and H168 and D172 for AT3-D1 and AT3-D2, respectively, are in magenta. Residues of the conserved motif DFGWG are in blue. **B** Protein surfaces are shown in AT3-D1 (left) and AT3-D2 (right). **C** Internal pocket of the AT3-D1 (left) and AT3-D2 (right). The internal cavity space is in brown, the conserved residues of histidine and aspartate of the motif HXXXD are represented as sticks. Figure **A** was made with PyMOL v.2.3.0. Figure **B** and **C** were made in the ProteinPlus platform.

4.3.5. Docking simulation with probable substrates

Docking analysis of the probable AT3 substrates are showed (Figure 4.4). First, the (E)-8-methylnon-6-enoyl-CoA docking was revised. In AT3-D1, the acyl moiety was oriented in an opposite side of the His169 and Asp173, which were oriented near to a phosphate group and the purine ring of the coenzyme A (Figure 4.4A). In AT3-D2, the acyl moiety was oriented nearest to the His168, which imidazole N_{ε2} is oriented to a carbonyl group of the coenzyme A, while the carbonyl group of the residue is oriented to the nitrogen of the amide group of Asn303 (Figure 4.4A). In general, the positioning of the acyl-CoA through the solvent channel allows a better approach between the acyl moiety carbonyl group and the N_{ε2} of the imidazole of the histidine in AT3-D2, although in both cases the distance was relatively high (8.9 Å and 7.9 Å for AT3-D1 and AT3-D2, respectively). The interaction energies were -8.6 kcal/mol and -9.3 kcal/mol in AT3-D1 and AT3-D2, respectively (Table 4.2). Thus, AT3-D2 model seems has a more favorable interaction with the (E)-8-methylnon-6-enoyl-CoA substratum. The 8-methylnonanoyl-CoA docking was revised. In AT3-D1, the carbonyl group of the 8-methylnonanoyl moiety was positioned between His169 and Asp173 side-chains (Figure 4.4B). In AT3-D2, the acyl moiety was surrounded mainly by non-polar and aromatic residues. The carbonyl group was positioned between His168 and Asp172, but the Asp172 side-chain was keep way from the substrate. The N_{ε2} of His 168 was oriented 4.1 Å to the carbonyl group of acyl moiety (Figure 4.4B). Interaction energies were quite similar being -8.6 kcal/mol and -8.4 kcal/mol for AT3-D1 and AT3-D2, respectively (Table 4.2). Vanillylamine docking was also revised. In AT3-D1, the amino group of the vanillylamine was positioned between the N_{ε2} of imidazole side-chain of His169 and the Asp173 oxygen carboxyl side-chain, in a position favorable for the interaction with the motif HXXXD in the catalytic site (Figure 4.4C). In AT3-D2, the vanillylamine amino group was keep away from the His168 side-chain, which maintains an orientation of its carbonyl group of the backbone with the amide group of the Asn303 side-chain. Asp172 side-chain is near of the guanidine side-chain of Arg289, which probably helps to

stabilize the conformation observed between these residues (Figure 4.4C). Interaction energies were quite similar with -5.1 kcal/mol and -5,0 kcal/mol for AT3-D1 and AT3-D2, respectively (Table 4.2). Even though, AT3-D1 protein model seems have a more favorable docking conformation, according to amino group orientation of the vanillylamine in the catalytic site with His169.

Previous phylogenetic analysis placed to AT3 encoded protein immediately basal to DAT (Stewart Jr *et al.*, 2005), an acyltransferase from *Catharantus roseus*, that is involved in the vindoline biosynthesis pathway and has as substrates the 4-O-deacetylvindoline and the acetyl-CoA (Laflamme *et al.*, 2001). Therefore, these substrates also were selected in the docking analysis with comparison aim. Acetyl-CoA docking was revised. In AT3-D1, the His169 and Asp173 was positioned in an opposite side of the acyl group of the acetyl-CoA, impossibility the interaction with the carbonyl group of the acyl moiety. In AT3-D2, the Asp172 side-chain carboxyl group also had a possible interaction with the guanidine group of Arg289, while its amino group of the backbone interacts with the carbonyl group of the Gly302 backbone whereas than the His168 residue was away from the acetyl group of the ligand, unable to interact with it. Interaction energies were -8.3 kcal/mol and -8.9 kcal/mol in AT3-D1 and AT3-D2, respectively (Table 4.2). Despite the different of the energy interactions, in both models, the catalytic histidine of the motif HXXXD was keep way from the carbonyl group of the acetyl moiety, making hard the reaction model of the acyltransferases (Supplementary Figure 4.3A). The 4-O-deacetylvindoline docking was revised. This substratum has a 4-hydroxyl group as the acyl acceptor. In the AT3-D1 model, the 4-hydroxyl group was positioned away from the His169 and Asp173 (6.5 Å and 8.7 Å, respectively). In AT3-D2, the 4-hydroxyl group was positioned facing to the His168 (4.0 Å) whereas than the Asp172 was away from the ligand. Energy interactions were -8.2 kcal/mol and -7.0 kcal/mol for AT3-D1 and AT3-D2, respectively (Table 4.2). In general, the position of the 4-hydroxyl group of the 4-O-deacetylvindoline acyl acceptor make difficult the expected acyltransferase reaction in both AT3 models (Supplementary Figure 4.3B).

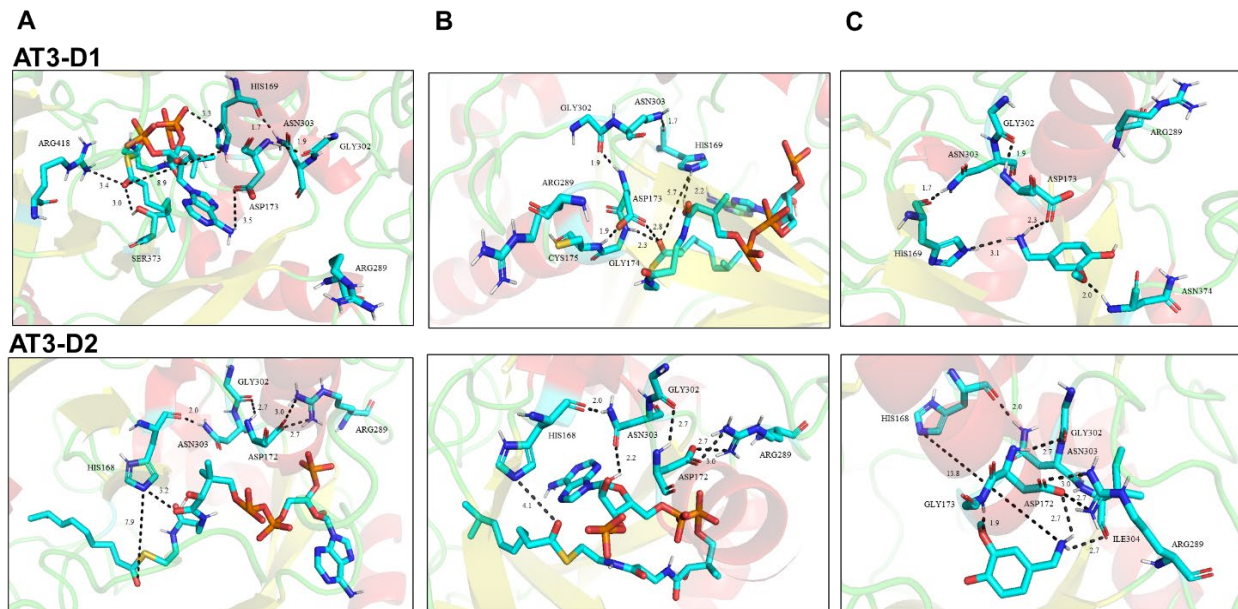


Figure 4.4. Docking results of AT3 models with the CS substrates **A** (E)-8-methylnon-6-enoyl-CoA, **B** 8-methylnonanoyl-CoA and **C** vanillylamine. Distances with nearest and catalytic residues are showed. AT3-D1 models (Up images), AT3-D2 models (Low images). Docking was made with Autodock Vina v.1.1.2. with AutodockTools v.1.5.6. Figures were made with PyMOL v2.3.0. Distances between atoms, in angstroms (Å), are signaled with a dotted line.

4.4. DISCUSSION

4.4.1. AT3 phylogenetic and haplotype analysis

Capsicum *AT3* sequences were selected according its reported function (Stellari *et al.*, 2010; Stewart Jr *et al.*, 2007; Stewart Jr *et al.*, 2005). All the sequences have 1 intron and 2 exons, and the BAHD acyltransferase domains. Some sequences from non-pungent accessions have a deletion in the exon 1 and, therefore, they were discarded. This trait was reported by other authors (Qin *et al.*, 2014; Stewart Jr *et al.*, 2005) as a common characteristic in the non-spicy accessions of *C. annuum*. Other accessions with premature stop codons as *cch05* (*pun1²* cultivar) also were discarded (Stewart Jr *et al.*, 2007). The sequence (*cfr05*) with mutations close to DFGWG motif, and previously reported as *pun1³*, was also discarded (Stellari *et al.*, 2010). *C. rhomboideum* *crh01* and *crh02* sequences were also removed to further analysis since they are pseudogenes copies which were previously identified by Egan *et al.* (2019). A few sequences (*can17*, *can21* and *can24*) were closely grouped with the 3 β -hydroxysteroid-dehydrogenase/descarboxylase also were discarded. Finally, thirty-two *AT3* sequences

covering *Capsicum* pungent and non-pungent accessions were identified as probable and functional CS *Capsicum*. AT3 protein sequences were grouped in two subclades (AT3-D1 and AT3-D2). Only two sequences (can14 and can16) in the AT3-D2 group were slight grouped to hypothetical DAT and therefore, these were included in this analysis. First phylogenetic analysis placed to AT3 encoded protein immediately basal to DAT (Stewart Jr *et al.*, 2005), an acyltransferase from *C. roseus* that is involved in the pathway vindoline biosynthesis (Laflamme *et al.*, 2001). The sequences of the groups AT3-D1 and AT3-D2 were reported by Qin *et al.* (2014), who analyzed the expression of various AT3 copies in pungent and non-pungent *C. annuum* accessions. AT3-D1 and AT3-D2, equivalents to the groups named AT3-D1 and AT3-D2 in the current study, exhibits a differential and low expression in the non-pungent accessions, reason why the authors conclude a dosage compensation effect between these AT3 copies. Occurrence of different functional AT3 copies in tandem in *Capsicum* and other Solanaceae plants also was reported by Kim *et al.* (2014) and Egan *et al.* (2019) where were found several copies of acyltransferases in the region of chromosome 2 suggesting that the capsaicin synthase gene and its paralogues were originated by five rounds of unequal duplication events from two ancestral paralogues, and in the last duplication event, the capsaicin synthase gene emerged in the *Capsicum* genome (Kim *et al.*, 2014). The gene AT3 has been reported in several works as one of the genes that determine the presence or absence of capsaicinoid in *Capsicum* fruits, but as far as we know, generally the gene has been studied without regard in the several copies that can exist in the *Capsicum* genomes (Tanaka *et al.*, 2017; Arce-Rodríguez and Ochoa-Alejo, 2015; Stellari *et al.*, 2010; Stewart Jr. *et al.*, 2007; Stewart Jr. *et al.*, 2005).

4.4.2. Regulatory elements

AT3-D1 and AT3-D2 gene groups were analyzed in the promoter region. A previous study has identified polymorphisms in this region among different *C. annuum* accessions (19 of 36 polymorphism) suggesting an effect on the levels of capsaicin and dihydrocapsaicin in these accessions (Reddy *et al.*, 2014)). AT3-D1 and AT3-D2 copies are differentially expressed in *Capsicum* cultivars (Qin *et al.*, 2014), but most of studies no characterize individually these different copies. In the current study, abscisic acid, abiotic stress and light are important factors that control the AT3 expression in *Capsicum*. Similar regulatory elements were reported in the promoter of AT3 gene in *C. annuum* CM334, where stimuli as the light, heat stress and the mechanical wounding were identified (Kim *et al.*, 2009). Upstream regions in AT3-D1 and AT3-D2 exhibit differences in the mechanical wounding factors that suggest a differential

transcriptional response to this stimulus. Mechanical wounding has also been found to increase the *AT3* expression in *Capsicum* (Kirke *et al.*, 2018; Kim *et al.*, 2009). Upstream regions in *AT3-D1* and *AT3-D2* also exhibit differences in the light response elements, which could indicate a difference in the response to light between them. Previous study has showed that *AT3* gene expression is higher under light conditions in comparison to dark or dim lighting conditions, which suggests that the light activates the *AT3* expression (Kim *et al.*, 2009). An MBS transcription factor for response to drought was found in *AT3-D1* and in *AT3-D2* of *C. chinense* (cch03) but no in *AT3-D2* of *C. annuum* (can06), which suggest that this abiotic factor plays a role in the differential expression in the *AT3* copies in different species. *C. chinense* can have a high content of capsaicinoids in response to water deficits conditions with an increase in CS activity (Zamudio-Moreno *et al.*, 2014), and the results suggests that both, *AT3-D1* and *AT3-D2* have a similar response to drought that could influence in the high capsaicinoid response in *C. chinense* in these conditions. MYB factor was the most abundant in the *AT3* copies. MYB transcription factors controls the gene expression in the capsaicinoid metabolism (Sun *et al.*, 2020; Han *et al.*, 2019; Arce-Rodríguez and Ochoa-Alejo, 2017). Regulatory regions of many capsaicin associated genes contain MYB binding (Zhu *et al.*, 2019). Specifically, MYB transcription factor CaMYB48 and CaMYB31 has been reported in the capsaicinoid regulation in *C. annuum* (Sun *et al.*, 2020; Arce-Rodríguez and Ochoa-Alejo, 2017) and the silencing of CaMYB31 affects the capsaicinoid metabolism (Arce-Rodríguez and Ochoa-Alejo, 2017). Similar evidence of MYB genes controlling the pungency and the expression of genes of the capsaicinoid biosynthetic pathway was found by Han *et al.* (2019). Similarly, MYB transcription factor CaMYB108 shows evidence of capsaicinoid content and floral development regulation in *C. annuum* (Sun *et al.*, 2019). In the current study, a F-box for floral development was found in *AT3-D1* but no in *AT3-D2* in both species. There has been evidence that the promoter of *Pun1* drivers a higher expression preference in reproductive organs like ovary, anthers and pollen, and fruits than in vegetative organs (Kirke *et al.*, 2018). Also, the cells where the capsaicinoids are synthesized and accumulated have a specific morphology and spatial distribution in the epidermis of the placental septum of the pungent fruits (Tanaka *et al.*, 2021). The presence of regulatory elements involved in the floral development and the existence of capsaicinoid biosynthetic cells at early anthesis suggests a *AT3* regulation in the floral stages and not only during fruit development. Recently, it was showed evidence that MYB31 control the transcription of genes implied in capsaicinoids biosynthetic, including the *AT3* gene, in the placenta and in the pericarp of highly pungent *C. chinense* cultivars (Sun *et al.*, 2022). Specifically, the cells where the capsaicinoids occurs are the placental septum and in the pericarp without a clear

difference in the boundaries between them, and this particular morphology beginning in the anthesis stage (Tanaka et al., 2021). Differences in the regulatory (F-box, light, MBS, MYB and fitohormones binding sites) factors between *AT3-D1* and *AT3-D2* copies suggests a differential regulation during floral development in *Capsicum*.

4.4.3. AT3 protein model

AT3 models obtained have C-scores ranked within -5 to 2 that suggests confident models (Yang and Zhang, 2015). TM-scores were also high (0 to 1) reflecting structural similarity with the template models (Zhang, 2008). In general, RMSD values were low with exception of AT3-D2 model and the HCT template (Hernández-Rodríguez *et al.*, 2016). AT3 protein structures are similar to the general structure of the acyltransferases of the superfamily BAHD with two domains formed each one by a group of β -strands. Both domains are connected with a large loop (Ma *et al.*, 2005). HKIGD domain has a conserved catalytic histidine and an aspartate that is believed is important for the substrate binding and for the correct conformation in active site in the solvent channel during catalysis (Manjasetty *et al.*, 2012). DFGWG motif is away from the active site and it is believed that have a structural role in favor of the enzyme activity (Ma *et al.*, 2005). Also, we found a GN sequence (Gly302 and Asn303) and a conserved Arg289 that has been reported conserved in some acyltransferases and that has been reported to have interactions with the motif HXXXD and it is believed that they favor the catalysis or the geometry of the active site (Walker *et al.*, 2013; Ma *et al.*, 2005). Finally, some polymorphic changes were observed close to the conserved motifs HKIGD and DFGWG between the AT3-D1 and AT3-D2 models. Similar aminoacid changes were reported by Qin *et al.* (2014).

A channel was found inside the protein, with the HXXXD in the central part of it and the side-chains of the conserved His169 and Asp173 of can01, and His168 of can12 exposed to the central region, but although Asp172 was part of the channel of can12, its side-chain keep away from that central region. Also, entrances to the channel were found in both sides of the models. This channel is called the solvent channel and is present in other acyltransferases, as in the template models used, vinorine synthase and HCT, and serves to the entrance of the substrates to the central pocket (Walker *et al.*, 2013; Ma *et al.*, 2005). The channel can differ in composition, shape and size in dependency of the specific acyltransferase, including members of the same acyltransferase family (Galaz *et al.*, 2013), and affect the union and affinity for different substrates (Galaz *et al.*, 2013; Morales-Quintana *et al.*, 2012). In this regard, the channel of can01 had a bigger volume and surface and minor depth than can12 model

(Supplementary Table 4.4). The shape of the channel of can01 model was central, crossing through the interdomain space from one side to the other of protein and with additional pocket spaces coming out to the central pocket. The channel of the can12 also crosses through the interdomain space of the protein, but presents more spaces coming out of the central pocket, in an intercrossed shape. Also, the entrances to the solvent channel in can12 were bigger than in can01. This bigger entrance and presence of more peripheral pockets to the central space would allow space to enter and bind large substrates as fatty-acid acyl-CoA esters to the solvent channel.

4.4.4. Docking analysis

In this study, a simulation of the coupling of the AT3 models obtained with the different substrates was carried out since the enzyme extracts purification of this plant are complicated by the large amount of endogenous capsaicinoids. In addition, an extreme insolubility of the recombinant AT3 protein and the lack of commercially available substrates of BAHD acyltransferases have also been reported (Ogawa *et al.*, 2015; Stewart Jr *et al.*, 2005; Beuerle and Pichersky, 2002). We select the (E)-8-methylnon-6-enoyl-CoA and 8-methylnonanoyl-CoA since they form the capsaicin and the dihydrocapsaicin, which are the first and second most abundant capsaicinoids in *Capsicum* fruits (Fujiwake *et al.*, 1980). The main difference in the docking between AT3-D1 and AT3-D2 was the position of the ligands in the solvent channel, its distances with respect to the histidine and aspartate of the conserved motif HXXXD and the position of the Asp172 of AT3-D2, away from the central pocket. In docking simulation of other acyltransferases models, the acyl-CoA enters to the central pocket positioning its acyl moiety nearest to the catalytic histidine, and the acyl acceptor is positioned nearest to the conserved aspartate of the HXXXD motif (Galaz *et al.*, 2013; Morales-Quintana *et al.*, 2012; Morales-Quintana *et al.*, 2011). This characteristic was observed in our models. In the current study, the acyl-CoA ligands were variable in its binding position in the solvent channel. In the case of AT3-D2, the carbonyl group of the acyl moiety was oriented nearest to the His168 for both (E)-8-methylnon-6-enoyl-CoA and 8-methylnonanoyl-CoA ligands in comparison with the equivalent His169 for AT3-D1. Besides, AT3-D1 model has an interaction energies differential between both ligands. This might be of biological significance since in cell free extracts of placenta has been observed different capsaicinoids contents when equivalent amounts of iso-acyl-CoA precursors are added, which could indicate a relation between the capsaicin synthase activity and the structure of the acyl-CoA (Fujiwake *et al.*, 1980). It is important to note that the pool of fatty acids and acyl-CoA precursors in the fruits could reflect the proportions of capsaicinoids,

so the proportion of each substrate is not necessary equivalent in nature and could direct the proportions of capsaicinoids formed in a different way that in cells extracts with the same proportion of each precursor (Thiele *et al.*, 2008; Fujiwake *et al.*, 1980). In addition, capsaicinoid biosynthesis not only depends on the AT3 activity, but also there are other important steps in the pathway, as the branched-chain fatty acid biosynthesis, in with proteins and genes as KAS, KR and FAT are associated with the capsaicinoid biosynthesis (Koeda *et al.*, 2019; Park *et al.*, 2019). The combination of the pool of acyl-CoA from the biosynthesis of branched-chain fatty acids and the activities of the AT3 with different acyl-CoA interaction might direct the preference in the formation of one specific capsaicinoid over other. Further studies are necessary to understand how all these enzymes work to direct the proportion of the different capsaicinoids observed in fruits in the nature.

Regarding to vanillylamine ligand, in AT3-D1, the amine group of this ligand was located near to the His169 but in AT3-D2 this amino group interacts with the carboxyl group of the Asp172 causing that this amino group keep away of the His168. This is important since the interaction of the acyl acceptor and the catalytic histidine is a key point in the acyltransferase reaction (Ma *et al.*, 2005). The interaction of the conserved aspartate with equivalents arginine residues has been described in some other acyltransferases such as vinorine synthase and HCT, suggesting that this aspartate could have a structural role in the binding site of the enzymes (Walker *et al.*, 2013; Ma *et al.*, 2005). By using molecular dynamic, it has been described that the aspartate is important in the binding of the acyl acceptor to later be reoriented to the catalytic histidine (Galaz *et al.*, 2013).

Hypothetical DAT sequences (KAF3623437 and KAF3637602) were slightly grouped with some few AT3-D2 sequences. DAT is involved in the vindoline biosynthesis and has as substrates to the 4-O-deacetylvindoline and to the acetyl-CoA (Laflamme *et al.*, 2001). Therefore, a further analysis was conducted with these probably DAT substrates to discard the DAT activity. The 4-O-deacetylvindoline ligand shows differential interactions of the His168 with the 4-hydroxyl group of the substrate, but the side-chain Asp172 was away from the central pocket. In the case of the acetyl-CoA ligand, the acetyl group was placed in the solvent channel in a conformation that makes hard their interaction in both models, specially, in AT3-D2 where the acetyl group was in the contrary side of the pocket. Together, both results could indicate that the AT3 models have not good DAT ligand interactions. Although, more efforts should be made to discard the DAT activity.

4.5. CONCLUSION

AT3-D1 and *AT3-D2* haplotypes were detected on *Capsicum* pungent and non-pungent accessions. These shows similar length CDS, conserved BAHD domains and exon numbers. However, some amino acid substitutions were found, particularly close to conserved HKIGD and DFGWG motifs. These haplotypes were localized on *C. annum* genome where were identified differential transcriptional factors on their upstream regions, mainly light and abiotic stress factors. Structure models of *AT3-D1* and *AT3-D2* results in a BAHD typical acyltransferase with high affinity to CS substrates. DAT activity was discarded by the distant interactions between the DAT substrates and the different binding sites of the *AT3* models. Furthermore, CS activity is supported by the high affinity between the vanillylamin and the acyl-CoA substrates with the different binding positions in the active site of the *AT3* models. Taken together these results, it is possible suggest that *AT3* sequences codes different CS copies with different transcriptional factors and different substrates affinities which occurs in all analyzed *Capsicum* species, and that might have a contribution in the regulation of capsaicin synthesis pathway.

CAPÍTULO V

5.1. DISCUSIÓN GENERAL

En este trabajo se comparó el efecto de suelos con diferentes características fisicoquímicas y microbiológicas (SN y SR) sobre la expresión de genes relacionados con el metabolismo de capsaicinoides en frutos de plantas de chile habanero. Se observó primeramente que hubo efecto sobre el contenido de capsaicinoides por el tipo de suelo que dependió del estado de desarrollo del fruto. Se encontró que a 25 DPA hubo una mayor acumulación de capsaicinoides en SN (10.72 mg/g PS de fruto) comparado con SR (8.92 mg/g PS de fruto), y luego, a 45 DPA el contenido de capsaicinoides en SN no se vio alterado, pero el de SR aumentó, de manera diferente al máximo alcanzado en SN.

Esto que indicó un favorecimiento hacia la síntesis y un bajo catabolismo de capsaicinoides en chile habanero crecido en los suelos negros (Figura 3.3). Esta estimulación en el contenido de capsaicinoides en SN podría deberse a factores como el contenido de nutrientes, el contenido de nitrógeno y la forma química en la que se encuentre (Zhang *et al.*, 2020a), especialmente como nitrato, que es indispensable para la síntesis de aminoácidos precursores y capsaicinoides en la placenta de chile habanero (Baas-Espinola *et al.*, 2016; Monforte-González *et al.*, 2010).

Aunque esto depende también del aprovechamiento real que tenga la planta, que está relacionado, además de la forma química del nutriente, con propiedades fisicoquímicas del suelo (textura, estructura, porosidad) (Schaeztl y Anderson, 2005), y capacidad de los nutrientes para salir de las partículas del suelo y movilizarse desde la solución del suelo hacia las raíces de la planta (Borges-Gómez *et al.*, 2005). Los suelos negros con alto contenido de materia orgánica y ácidos fúlvicos y húmicos también favorecen el contenido de capsaicinoides (Das *et al.*, 2016).

Una vez encontrado el comportamiento en la acumulación de capsaicinoides, se decidió usar los frutos de 25 DPA para realizar un análisis transcriptómico, debido a que en este tiempo de desarrollo se encontró una diferencia en el contenido de capsaicinoides, y a que se conoce que son estas etapas intermedias de desarrollo en las que el fruto es más activo en la acumulación de estos metabolitos (Benmansour-Gueddes, 2015), y donde la expresión de los genes relacionados a su metabolismo suele ser más alta (Martínez-López *et al.*, 2014).

Al analizar los genes más expresados en SN y las rutas metabólicas enriquecidas, muchos de estos estuvieron relacionados con la fotosíntesis. Principalmente tuvieron mayor expresión proteínas de unión a clorofila de los LHC, GluTR, GSA, RuBisCO, DXS y DXS2, y estuvieron enriquecidas las rutas de metabolismo de porfirinas, lo que indicó una posible estimulación de la fotosíntesis que podría favorecer el desarrollo del fruto y procesos de biosíntesis (Borovsky *et al.*, 2019; Lupi *et al.*, 2019; Nguyen *et al.*, 2014). Aunque no se observó una diferencia global en la producción de frutos (Figura 3.2A), hubo, en mayor proporción, frutos de 45 DPA de SN en estado de cambio de coloración de verde a naranja y menor proporción de frutos aún en coloración verde (Tabla Suplementaria 3.4).

En este rasgo, aunque en los frutos de 25 DPA no se observó diferencia de la expresión de PSY2 y PSD, si se observó diferencia de expresión en los ortólogos de DXS y DXS2, de la biosíntesis de terpenoides (precursores de la ruta de los carotenoides), y enriquecimiento de las rutas de síntesis de terpenoides y carotenoides, lo que significa que la síntesis de carotenoides a los 25 DPA podría estar empezando a estimularse, y ya a 45 DPA observarse el efecto favorecido en los frutos crecidos en SN (Martínez-López *et al.*, 2014).

Además, lo anterior, junto al enriquecimiento de rutas como las relacionadas con la síntesis de valina, leucina, isoleucina, arginina, glutamato, aspartato y alanina, la ruta de las pentosas fosfato, y la mayor expresión de genes como de albumina, *6GPDH*, *6PGD* y *KARI* indiquen también una mayor estimulación para la síntesis y el metabolismo del nitrógeno, aminoácidos esenciales, proteínas y azúcares (Figura 3.5, Figura 3.6 y Tabla Suplementaria 3.1), esto posiblemente debido al mayor contenido de nutrientes, materia orgánica y tipos de microorganismos de SN que lo diferencian de SR (Das *et al.*, 2016; Estrada-Medina *et al.*, 2016).

Cabe mencionar que esta expresión y rutas metabólicas enriquecidas se observó en frutos a los 25 DPA, por lo que su influencia en frutos de 45 DPA, que están entrando a madurez, aún debe comprobarse. Otros genes con mayor expresión en SN estuvieron relacionados con el desarrollo general de frutos y respuesta al medio ambiente. (Figura 3.5A). En particular, *BREVIS RADIX* puede influir en el desarrollo del número de lóculos de los frutos (Ma *et al.*, 2022). Sin embargo, los resultados obtenidos de la transcriptómica, el contenido de capsaicina a 25 DPA y la producción de frutos nos hacen creer que SN estimula, más que el crecimiento del fruto, procesos metabólicos que pueden ser favorables para el contenido nutricional de éstos. Por el contrario, SR, tuvo solo la ruta de las antocianinas enriquecida, y tuvo mayor

expresión de genes relacionados con el desarrollo floral, la fertilidad de semillas, desarrollo de frutos y de las paredes celulares (Figura 3.5). Esto puede indicar que SR estimule más procesos de desarrollo del fruto desde el desarrollo floral hasta la madurez, incluyendo el ablandamiento de los frutos (softening) y la abscisión, por lo que los suelos rojos serían más favorables para la producción (Leszczuk *et al.*, 2020; Zhang *et al.*, 2019; Ihsan *et al.*, 2015; Park *et al.*, 2015b). Esto puede explicar el hecho de que hubo más frutos de mayor tamaño en plantas crecidas en SR, aun cuando la producción global en ambos suelos fue similar, (Figura 3.2). Los resultados anteriores se resumen en la Figura 5.1A.

En los genes directamente relacionados con la ruta de síntesis de capsaicinoides no se observó diferencia de expresión en tipos de suelo. Sin embargo, sí se encontró mayor expresión en genes que pudieran estar relacionados con este proceso (Figura 3.6). Así, los genes de la ruta de las pentosas fosfato *G6PDH* y *6GPD* tuvieron mayor expresión en SN, y particularmente *G6PDH* estuvo en el mismo clado que la mayoría de los genes de la ruta de síntesis de capsaicinoides. Estos genes y la ruta de las pentosas fosfato podrían participar en la actividad antioxidante de los frutos (Palma *et al.*, 2020), y disminuir el catabolismo de los capsaicinoides por las POD (Zamudio-Moreno *et al.*, 2014), lo cual pudiera haber influenciado a que en SN a los 25 DPA se hubiera alcanzado un máximo en el contenido de capsaicinoides antes que SR, es decir, disminuyó la degradación de los capsaicinoides, y por lo tanto aumentó su acumulación en los frutos de SN (Figura 3.2).

También, estos genes podrían participar en la obtención de NADPH, que particularmente es importante en la síntesis de los grupos acilo que conforman a las capsaicinoides (Koeda *et al.*, 2019). El gen *RPT4* pudiera estar relacionado con el transporte vesicular y vacuolar en las células del septum placentario, y al ser los frutos de SN más activos en la acumulación de capsaicinoides, este proceso celular es esperado que sea más activo en comparación a los frutos de SR (Tyler *et al.*, 2016). Además, en frutos de variedades muy picantes de *C. chinense*, se ha observado la acumulación de capsaicinoides en sus pericarpios, así como cambios celulares en el pericarpio similares a los observados en las células de acumulación de capsaicinoides del septum placentario (Tanaka *et al.*, 2021), por lo que el aparato del proteasoma podría ser no solo importante en la acumulación de capsaicinoides en la placenta, sino también en pericarpio de esas variedades super picantes. *KARI* también fue más expresado en SN. Al ser *KARI* una enzima importante en la síntesis de valina y leucina, precursores de los capsaicinoides, su importancia en la mayor expresión de su gen puede estar

relacionada con una mayor síntesis de estos aminoácidos en las placentas de los frutos de SN (Baas-Espinola *et al.*, 2016).

Aunque los genes propios de la ruta de síntesis de capsaicinoides no tuvieron diferencia en su expresión a 25 DPA entre suelos, esto no significa necesariamente el mismo nivel de actividad de las proteínas codificadas por dichos genes. Para profundizar en la afinidad que pueden tener enzimas de la ruta de síntesis de capsaicinoides sobre sus sustratos, se seleccionó a AT3 al ser una enzima de un punto de control clave en la ruta de síntesis de capsaicinoides (Arce-Rodríguez y Ochoa-Alejo, 2019; Ogawa *et al.*, 2015).

El análisis genómico previo en *Capsicum* mostró dos grupos principales de haplotipos: AT3-D1 y AT3-D2 (Figuras 4.1 y 4.2). Estos también correspondieron a los dos ortólogos de AT3 a los que se les analizó su expresión en los frutos de chile habanero, teniendo AT3-D1 mayor expresión que AT3-D2 en ambos tipos de suelo (Figura 3.6), lo cual se ha reportado por otros autores, siendo AT3-D1, al menos en expresión de gen, el principal haplotipo relacionado con la síntesis de capsaicinoides (Qin *et al.*, 2014).

El análisis de elementos regulatorios en ambos haplotipos mostró diferencias en los elementos relacionados para estímulos como daño mecánico, luz, sequía, así como relacionados al desarrollo floral, la síntesis de flavonoides, el ciclo circadiano y estrés abiótico en general (Tabla 4.1). Muchos de estos estímulos se han reportado que afectan el contenido de capsaicinoides y a la expresión de AT3 (Tanaka *et al.*, 2021; Sun *et al.*, 2020; Kirke *et al.*, 2018; Zamudio-Moreno *et al.*, 2014; Kim *et al.*, 2009). Sin embargo, este análisis reveló que además ambos haplotipos responden de manera diferente a dichos estímulos, lo cual es importante para controlar la síntesis de capsaicinoides si cada haplotipo tiene un aporte distinto en la actividad de capsaicina sintasa de los frutos (Qin *et al.*, 2014).

Los modelos proteicos de AT3 obtenidos mostraron una estructura similar a las aciltransferasas de la familia superfamilia BAHD, lo cual coincide con la predicción por filogenia de sus genes (Stewart Jr *et al.*, 2005), con motivo catalítico HKIGD conservado en el bolsillo de unión de sustrato característico, y el motivo DFGWG en una estructura en forma de lazo alejada del sitio activo, también característica de este grupo de proteínas (Figura 4.3) (Ma *et al.*, 2005). Las estructuras de AT3-D1 y AT3-D2 difirieron en el canal del solvente donde se encuentra el sitio de unión de sustratos, teniendo AT3-D1 un canal con un volumen y superficie mayores que AT3-D2, y una forma central que cruzó de un extremo a otro la estructura de la proteína a lo

largo del espacio entre los dos dominios, con espacios adicionales saliendo del bolsillo central. Aunque AT3-D2 tuvo menor volumen y superficie, tuvo más espacios adicionales que partieron del dominio central, y una entrada al canal más grande que AT3-D1 (Figura 4.3B y 4.3C).

Estas diferencias son relevantes, porque sugieren una diferencia de afinidad por los sustratos (Galaz *et al.*, 2013). El análisis de docking también mostró diferencias en la afinidad por la (E)-8-metilnon-6-enoil-CoA, la 8-metilnonanoil-CoA y la vanillilamina, sustratos propuestos para la capsaicina sintasa (Ogawa *et al.*, 2015), difiriendo la posición de los ligandos en el sitio de unión, y la posición de los residuos de histidina y de aspartato del motivo catalítico HKIGD en el sitio de unión y con respecto al sustrato, y con AT3-D2 mostrando mayor diferencia de energía de afinidad entre los sustratos de acil-CoA que lo observado en AT3-D1 (Figura 4.4 y Tabla 4.2).

Esto puede indicar que cada haplotipo de AT3 contribuye de manera distinta en la síntesis de los diferentes capsaicinoides, como la capsaicina y la dihidrocapsaicina, para finalmente dirigir el contenido de capsaicinoides hacia las proporciones que se encuentran naturalmente (Benmansour-Gueddes, 2015). Sin embargo, esto no sólo depende de la estructura del sustrato, y de la capsaicina sintasa, sino también del contenido y proporción de precursores acil-CoA disponibles (Thiele *et al.*, 2008; Fujiwake *et al.*, 1980) y los genes y enzimas de la síntesis de ácidos grasos (Koeda *et al.*, 2019), por lo que también sería importante el estudio de la afinidad de KAS, FAT y KR en la síntesis de los acil-CoA que forman los capsaicinoides. En la Figura 5.1B se resumen los resultados hallados de la expresión de genes en los frutos de SN y SR, y de los análisis de docking de los haplotipos de AT3.

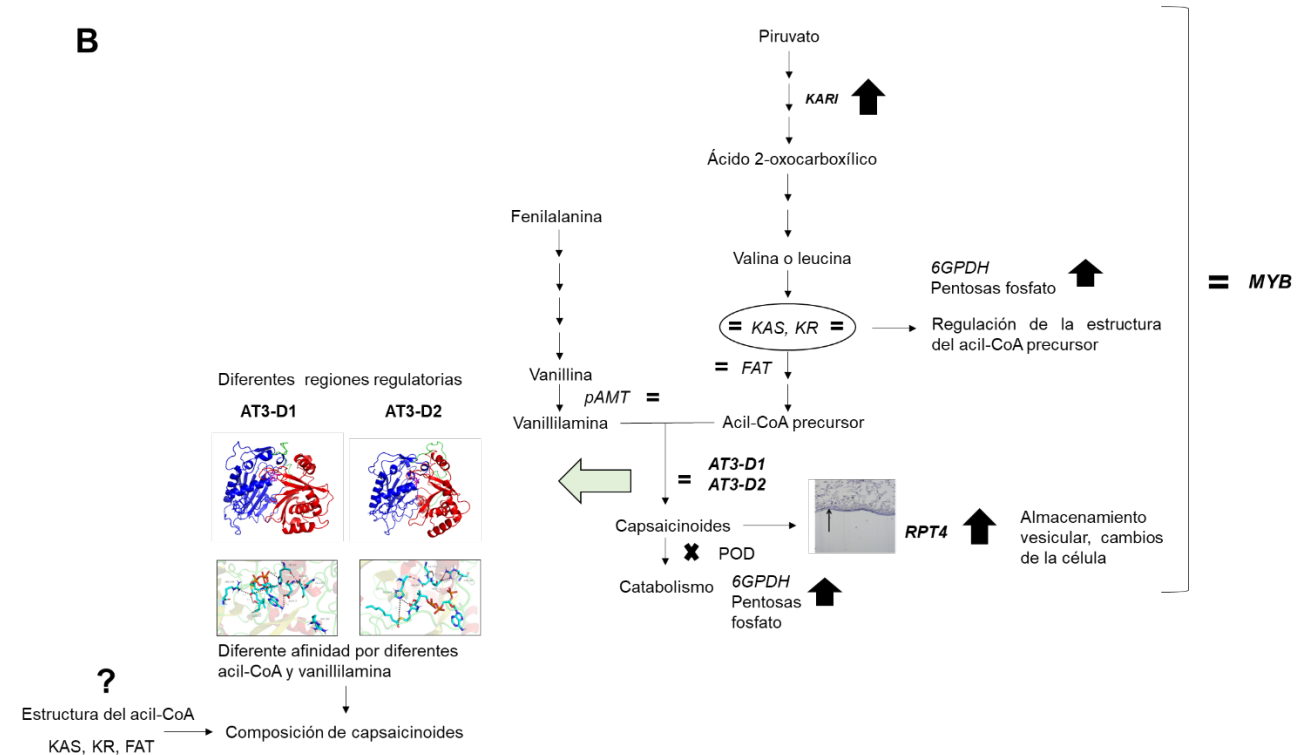
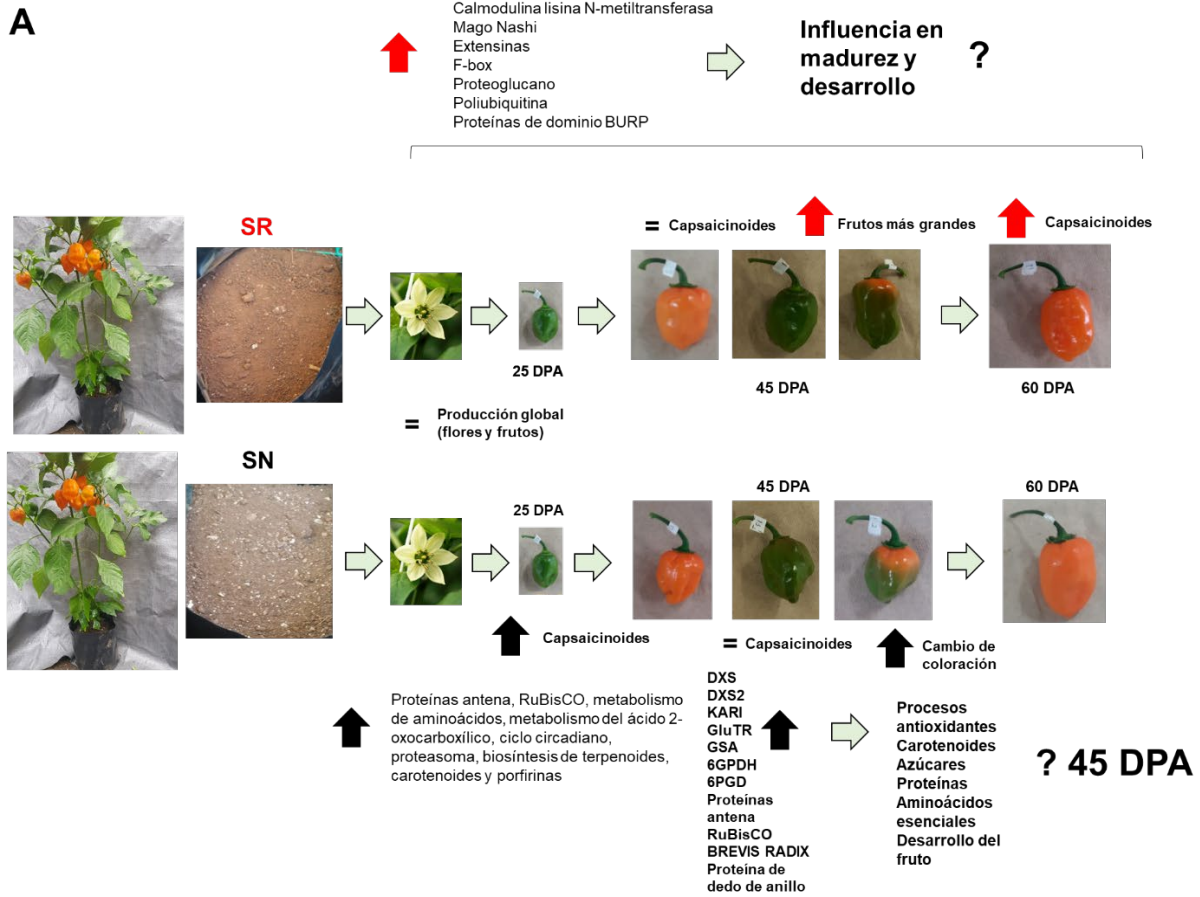


Figura 5.1. Resumen general de los principales hallazgos de este trabajo. A) Efecto del tipo de suelo (SN o SR) sobre los parámetros fisiológicos y el contenido de capsaicinoides en las plantas y frutos de chile habanero, así como en las rutas metabólicas enriquecidas y la expresión de genes relacionados con la producción. B) Efecto del tipo de suelo sobre los genes relacionados con la síntesis de capsaicinoides en frutos de chile habanero de 25 DPA y relación con los haplotipos de *AT3* y la afinidad de sus modelos proteicos hacia los sustratos de la capsaicina sintasa. Con una flecha negra grande se muestran los efectos que fueron mayores en SN, con una flecha roja los efectos que fueron mayores en SR, y con un signo de igual (=) aquellos que no fueron diferentes entre suelos

5.2. CONCLUSIONES GENERALES

Las plantas de chile habanero tuvieron una mayor producción de frutos más desarrollados de 45 DPA en SR que en SN. Sin embargo, la producción total de frutos no fue diferente entre tipos de suelo. El contenido de capsaicinoides dependió tanto del tipo de suelo como del estado de desarrollo de la planta, alcanzándose un mayor contenido de capsaicinoides en SN en frutos de estado de madurez intermedio (25 DPA), que sólo disminuyó tras los 45 DPA. El transcriptoma entre tipos de suelo también mostró cambios relacionados con la producción entre tipos de suelo. En SN hubo mayor número de rutas metabólicas enriquecidas y mayor expresión de genes relacionados con fotosíntesis, clorofila, terpenoides, ruta de las pentosas fosfato, proteasoma y síntesis de aminoácidos, mientras que SR tuvo una mayor expresión de genes relacionados con el desarrollo floral y la madurez del fruto. Aunque no hubo diferencia en la expresión de los genes directamente relacionados con la síntesis de capsaicinoides a 25 DPA, genes de rutas relacionadas (como *KARI* de la síntesis de aminoácidos de cadena ramificada, *RPT4* del proteasoma y *6GPDH* de la ruta de las pentosas fosfato) fueron más expresados en SN. Esto sugiere una importancia del tipo de suelo en la contribución de rutas metabólicas indirectamente relacionadas con la síntesis de capsaicinoides y su acumulación en frutos. Dentro de la ruta de síntesis de capsaicinoides, *AT3* demostró dos grupos de haplotipos con diferente expresión. Éstos difirieron en las regiones reguladoras relacionadas principalmente con luz y estrés abiótico. La estructura proteica de ambos haplotipos fue correspondiente a la de las aciltransferasas de la superfamilia BAHD. Cada haplotipo difirió en su afinidad por diferentes sustratos acil-CoA, y diferentes posiciones de los sustratos de la CS en el bolsillo de unión. Las diferencias en las regiones reguladoras, estructura proteica y afinidad por sustratos de la CS sugieren diferente contribución de los AT3 en la síntesis de capsaicinoides.

5.3. PERSPECTIVAS

Se plantea comprobar la expresión de los genes relacionados estudiados en este trabajo mediante PCR Tiempo Real, y analizar la expresión en otros estados de desarrollo del fruto, como 45 y 60 DPA.

Asimismo, estudiar el efecto del tipo de suelo junto a otro estímulo abiótico, como la luz, la aplicación de fertilizantes o nutrientes o el estrés hídrico sobre el contenido de capsaicinoides y la expresión de genes relacionados, y si con esto el contenido de capsaicinoides podría aumentar sobre el obtenido en el presente trabajo. Particularmente, se plantea estudiar si existe una diferente respuesta en la expresión de *AT3-D1* y *AT3-D2* bajo efecto de un estímulo abiótico, y la relación de esta respuesta de los diferentes haplotipos en la síntesis de capsaicinoides.

También se plantea la posibilidad de profundizar en la contribución de los diferentes haplotipos de la AT3 continuando los estudios de acoplamiento molecular sobre los sustratos de la CS mediante docking flexible y dinámica molecular, y comprobar si cambian para aumentar las afinidades de los modelos proteicos.

BIBLIOGRAFÍA

- Agrawal, S., Karcher, D., Ruf, S., Bock, R. (2020). The functions of chloroplast glutamyl-tRNA in translation and tetrapyrrole biosynthesis. *Plant Physiology*, 183(1), 263–276. <https://doi.org/10.1104/pp.20.00009>
- Akhtar, A., Asghar, W., Khalid, N. (2021). Phytochemical constituents and biological properties of domesticated *Capsicum* species: a review. *Bioactive Compounds in Health and Disease*, 4(9), 201–225. <https://doi.org/10.31989/bchd.v4i9.837>
- Altúzar-Molina, A. R., Muñoz-Sánchez, J. A., Vázquez-Flota, F., Monforte-González, M., Racagni-Di Palma, G., Hernández-Sotomayor, S. M. T. (2011). Phospholipidic signaling and vanillin production in response to salicylic acid and methyl jasmonate in *Capsicum chinense* J. cells. *Plant Physiology and Biochemistry*, 49(2), 151–158. <https://doi.org/10.1016/j.plaphy.2010.11.005>
- Ananthan, R., Subhash, K., Longvah, T. (2018). Capsaicinoids, amino acid and fatty acid profiles in different fruit components Naga king chilli (*Capsicum chinense* Jacq). *Food Chemistry*, 238, 51–57. <https://doi.org/10.1016/j.foodchem.2016.12.073>
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Antonio, A. S., Wiedemann, L. S. M., Veiga Junior, V. F. (2018). The genus *Capsicum*: a phytochemical review of bioactive secondary metabolites. *RSC Advances*, 8(45), 25767–25784. <https://doi.org/10.1039/c8ra02067a>
- Antonious, G. F., Berke, T., Jarret, R. L. (2009). Pungency in *Capsicum chinense*: Variation among countries of origin. *Journal of Environmental Science and Health - Part B*, 44(2), 179–184. <https://doi.org/10.1080/03601230802599118>

- Aranha, B. C., Hoffmann, J. F., Barbieri, R. L., Rombaldi, C. V., Chaves, F. C. (2017). Untargeted Metabolomic Analysis of *Capsicum* spp. by GC–MS. *Phytochemical Analysis*, 28(5), 439–447. <https://doi.org/10.1002/pca.2692>
- Arce-Rodríguez, M. L., Ochoa-Alejo, N. (2015). Silencing *AT3* gene reduces the expression of *pAmt*, *BCAT*, *Kas*, and *Acl* genes involved in capsaicinoid biosynthesis in chili pepper fruits. *Biologia Plantarum*, 59(3), 477–484. <https://doi.org/10.1007/s10535-015-0525-y>
- Arce-Rodríguez, M. L., Ochoa-Alejo, N. (2017). An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiology*, 174(3), 1359–1370. <https://doi.org/10.1104/pp.17.00506>
- Arce-Rodríguez, M. L., Ochoa-Alejo, N. (2019). Biochemistry and molecular biology of capsaicinoid biosynthesis: recent advances and perspectives. *Plant Cell Reports*, 38(9), 1017–1030. <https://doi.org/10.1007/s00299-019-02406-0>
- Ávalos García, A., Pérez-Urria Carril, E. (2009). Metabolismo secundario de plantas. *Reduca (Biología)*. *Serie Fisiología Vegetal*, 2(3), 119–145. ISSN: 1989-3620
- Aza-González, C., Núñez-Palenius, H. G., Ochoa-Alejo, N. (2011). Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). *Plant Cell Reports*, 30(5), 695–706. <https://doi.org/10.1007/s00299-010-0968-8>
- Baas-Espinola, F. M., Castro-Concha, L. A., Vázquez-Flota, F. A., Miranda-Ham, M. L. (2016). Capsaicin synthesis requires in situ phenylalanine and valine formation in *in vitro* maintained placentas from *Capsicum chinense*. *Molecules*, 21(6), 799. <https://doi.org/10.3390/molecules21060799>
- Baenas, N., Belović, M., Ilic, N., Moreno, D. A., García-Viguera, C. (2019). Industrial use of pepper (*Capsicum annum* L.) derived products: Technological benefits and biological advantages.

Food Chemistry, 274, 872–885. <https://doi.org/10.1016/j.foodchem.2018.09.047>

Barboza, G. E., García, C. C., González, S. L., Scaldaferrro, M., Reyes, X. (2019). Four new species of *Capsicum* (Solanaceae) from the tropical Andes and an update on the phylogeny of the genus. *PLoS ONE*, 14(1), 1–26. <https://doi.org/10.1371/journal.pone.0209792>

Bautista-Zúñiga, F., Jiménez-Osornio, J., Navarro-Alberto, J., Manu, A., Lozano, R. (2003). MICRORELIEVE Y COLOR DEL SUELO COMO PROPIEDADES DE DIAGNÓSTICO EN LEPTOSOLES CÁRSTICOS. *Terra Latinoamericana*, 21(1), 1–11. ISSN: Disponible en: <https://www.redalyc.org/articulo.oa?id=57321101>

Bautista, F., Díaz-Garrido, S., Castillo-González, M., Zinck, J. A. (2005). Spatial Heterogeneity of the Soil Cover in the Yucatán Karst: Comparison of Mayan, WRB, and Numerical Classifications. *Eurasian Soil Science*, 38(suppl. 1), S81–S88.

Ben-Chaim, A., Borovsky, Y., Falise, M., Mazourek, M., Kang, B. C., Paran, I., Jahn, M. (2006). QTL analysis for capsaicinoid content in *Capsicum*. *Theoretical and Applied Genetics*, 113(8), 1481–1490. <https://doi.org/10.1007/s00122-006-0395-y>

Benmansour-Gueddes, S. (2015). Evolution of Capsaicinoids and Mineral Composition During Fruit Development in Some Hot Pepper Varieties (*Capsicum annum* L.). *International Journal of Agriculture Innovations and Research*, 3(4), 1094-1102.

Beuerle, T., Pichersky, E. (2002). Enzymatic synthesis and purification of aromatic coenzyme A esters. *Analytical Biochemistry*, 302(2), 305–312. <https://doi.org/10.1006/abio.2001.5574>

Bogusz Junior, S., Marçõ, P. H., Valderrama, P., Damasceno, F. C., Aranda, M. S., Zini, C. A., Caramão, E. B., Tavares Melo, A. M., Filho, J. T., Godoy, H. T. (2015). Analysis of volatile compounds in *Capsicum* spp. by headspace solid-phase microextraction and GC × GC-TOFMS. *Analytical Methods*, 7(2), 521–529. <https://doi.org/10.1039/c4ay01455c>

- Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borges-Gómez, L., Escamilla-Bencomo, A., Soria-Fregoso, M., Casanova-Villareal, V. (2005). Potasio en suelos de Yucatán. *Terra Latinoamericana*, 23(4), 437–445. <http://www.redalyc.org/articulo.oa?id=57311146002>
- Borges-Gómez, L., Moo-Kauil, C., Ruíz-Novelo, J. (2014). SUELOS DESTINADOS A LA PRODUCCIÓN DE CHILE HABANERO EN YUCATÁN: CARACTERÍSTICAS FÍSICAS Y QUÍMICAS PREDOMINANTES. *Agrociencia*, 48(4), 347–359. Disponible en: <https://www.redalyc.org/articulo.oa?id=30232499001>
- Borovsky, Y., Monsonego, N., Mohan, V., Shabtai, S., Kamara, I., Faigenboim, A., Hill, T., Chen, S., Stoffel, K., Van Deynze, A., Paran, I. (2019). The zinc-finger transcription factor *CcLOL1* controls chloroplast development and immature pepper fruit color in *Capsicum chinense* and its function is conserved in tomato. *Plant Journal*, 99(1), 41–55. <https://doi.org/10.1111/tbj.14305>
- Bosland, P. W., Coon, D., Cooke, P. H. (2015). Novel formation of ectopic (Nonplacental) capsaicinoid secreting vesicles on fruit walls explains the morphological mechanism for super-hot chile peppers. *Journal of the American Society for Horticultural Science*, 140(3), 253–256. <https://doi.org/10.21273/jashs.140.3.253>
- Bosland, P. W., Votava, E. J. (2012). Taxonomy, pod types and genetic resources, en: Peppers: vegetable and spices *Capsicum* (2nd ed.), Atherton, J. (Ed.). CABI, Wallingford, UK, pp. 13–23. DOI: 10.1079/9781845938253.0000
- Buchfink, B., Reuter, K., Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18, 366–368. <https://doi.org/10.1038/s41592-021-01101-x>

BIBLIOGRAFÍA

- Burgos-Valencia, E., Echevarría-Machado, I., Narváez-Zapata, J. A., Martínez-Estévez, M. (2020). Gene expression related to the capsaicinoids biosynthesis in the *Capsicum* genus: Molecular and transcriptomic studies. *Revista Brasileira de Botânica*, 43(1), 201–212. <https://doi.org/10.1007/s40415-019-00575-6>
- Calzada-Rovirosa, J. E., Narváez-Narváez, J. A., Aguilar-Castillo, R., Romero-Celis, M., López Sánchez, M., Flores-Lui, L. F., Ornelas-Ruiz, P., Sánchez-Cruz, E., Valencia-Abundis, M. S., Rodríguez-Santiago, B., Gutiérrez-Alonso, O., Díaz-Vázquez, A. (2017). Paquete tecnológico para la producción de chile habanero. In *Agenda Técnica Agrícola. Yucatán* (pp. 49–59). INIFAP, SAGARPA, COFUPRO.
- Canto-Flick, A., Balam-Uc, E., Bello-Bello, J., Lecona-Guzmán, C., Solís-Marroquín, D., Avilés-Viñas, S., Gómez-Uc, E., López-Puc, G., Santana-Buzzy, N. (2008). Capsaicinoids Content in Habanero Pepper (*Capsicum chinense* Jacq.): Hottest Known Cultivars. *HORTSCIENCE*, 43(5), 1344–1349. <https://doi.org/10.21273/HORTSCI.43.5.1344>
- Carrizo García, C., Barfuss, M. H. J., Sehr, E. M., Barboza, G. E., Samuel, R., Moscone, E. A., Ehrendorfer, F. (2016). Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Annals of Botany*, 118(1), 35–51. <https://doi.org/10.1093/aob/mcw079>
- Castro-Concha, L. A., Baas-Espinola, F. M., Ancona-Escalante, W. R., Vázquez-Flota, F. A., Miranda-Ham, M. L. (2016). Phenylalanine biosynthesis and its relationship to accumulation of capsaicinoids during *Capsicum chinense* fruit development. *Biologia Plantarum*, 60(3), 579–584. <https://doi.org/10.1007/s10535-016-0608-4>
- Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature*, 389(6653), 816–824. <https://doi.org/10.1038/39807>
- Chel-Guerrero, L. D., Oney-Montalvo, J. E., Rodríguez-Buenfil, I. M. (2021). Phytochemical

Characterization of By-Products of Habanero. *Plants*, 10(4), 779. <https://doi.org/10.3390/plants10040779>

Chen, Y., Zhou, B., Li, J., Tang, H., Tang, J., Yang, Z. (2018). Formation and change of chloroplast-located plant metabolites in response to light conditions. *International Journal of Molecular Sciences*, 19(3), 654. <https://doi.org/10.3390/ijms19030654>

Chhapekar, S. S., Brahma, V., Rawoof, A., Kumar, N., Gaur, R., Jaiswal, V., Kumar, A., Yadava, S. K., Kumar, R., Sharma, V., Babu, S. S., Ramchiary, N. (2020). Transcriptome profiling, simple sequence repeat markers development and genetic diversity analysis of potential industrial crops *Capsicum chinense* and *C. frutescens* of Northeast India. *Industrial Crops and Products*, 154, 112687. <https://doi.org/10.1016/j.indcrop.2020.112687>

Clark, R., Lee, S.-H. (2016). Anticancer properties of capsaicin against human cancer. *Anticancer Research*, 36, 837–844. PMID: 26976969

Curry, J., Aluru, M., Mendoza, M., Nevarez, J., Melendrez, M., O'Connell, M. A. (1999). Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Science*, 148(1), 47–57. [https://doi.org/10.1016/S0168-9452\(99\)00118-1](https://doi.org/10.1016/S0168-9452(99)00118-1)

Das, S., Sarkar, S., Das, M., Banik, P., Bhattacharya, S. S. (2021). Influence of soil quality factors on capsaicin biosynthesis, pungency, yield, and produce quality of chili: An insight on *Csy1*, *Pun1*, and *Pun1²* signaling responses. *Plant Physiology and Biochemistry*, 166, 427–436. <https://doi.org/10.1016/j.plaphy.2021.06.012>

Das, S., Teja, K. C., Duary, B., Agrawal, P. K., Bhattacharya, S. S. (2016). Impact of nutrient management, soil type and location on the accumulation of capsaicin in *Capsicum chinense* (Jacq.): One of the hottest chili in the world. *Scientia Horticulturae*, 213, 354–366. <https://doi.org/10.1016/j.scienta.2016.10.041>

BIBLIOGRAFÍA

- Di Sotto, A., Vecchiato, M., Abete, L., Toniolo, C., Giusti, A. M., Mannina, L., Locatelli, M., Nicoletti, M., Di Giacomo, S. (2018). *Capsicum annuum* L. var. Cornetto di Pontecorvo PDO: Polyphenolic profile and *in vitro* biological activities. *Journal of Functional Foods*, 40, 679–691. <https://doi.org/10.1016/j.jff.2017.11.041>
- Egan, A. N., Moore, S., Stellari, G. M., Kang, B. C., Jahn, M. M. (2019). Tandem gene duplication and recombination at the *AT3* locus in the Solanaceae, a gene essential for capsaicinoid biosynthesis in *Capsicum*. *PLoS ONE*, 14(1), 1–24. <https://doi.org/10.1371/journal.pone.0210510>
- Eggink, P. M., Maliepaard, C., Tikunov, Y., Haanstra, J. P. W., Bovy, A. G., Visser, R. G. F. (2012). A taste of sweet pepper: Volatile and non-volatile chemical composition of fresh sweet pepper (*Capsicum annuum*) in relation to sensory evaluation of taste. *Food Chemistry*, 132(1), 301–310. <https://doi.org/10.1016/j.foodchem.2011.10.081>
- Estrada-Medina, H., Bautista, F., Jiménez-Osornio, J. J. M., González-Iturbe, J. A., Aguilar Cordero, W. de J. (2013). Maya and WRB Soil Classification in Yucatan, Mexico: Differences and Similarities. *ISRN Soil Science*, 2013, 634260. <https://doi.org/10.1155/2013/634260>
- Estrada-Medina, H., Canto-Canché, B. B., De los Santos-Briones, C., O'Connor-Sánchez, A. (2016). Yucatán in black and red: Linking edaphic analysis and pyrosequencing-based assessment of bacterial and fungal community structures in the two main kinds of soil of Yucatán State. *Microbiological Research*, 188–189, 23–33. <https://doi.org/10.1016/j.micres.2016.04.007>
- FAO. (2015). World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. FAO. <http://www.fao.org>
- Ferreira, L. G., Dos Santos, R. N., Oliva, G., Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, 20(7), 13384-13421. <https://doi.org/10.3390/molecules200713384>

- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Fujiwake, H., Suzuki, T., Oka, S., Iwai, K. (1980). Enzymatic formation of capsaicinoid from vanillylamine and iso-type fatty acids by cell-free extracts of *Capsicum annuum* var. *annuum* cv. Karayatsubusaf. *Agricultural and Biological Chemistry*, 44(12), 2907–2912. <https://doi.org/10.1080/00021369.1980.10864424>
- Galaz, S., Morales-Quintana, L., Moya-León, M. A., Herrera, R. (2013). Structural analysis of the alcohol acyltransferase protein family from *Cucumis melo* shows that enzyme activity depends on an essential solvent channel. *FEBS Journal*, 280(5), 1344–1357. <https://doi.org/10.1111/febs.12127>
- Gangadhar, B. H., Mishra, R. K., Pandian, G., Park, S. W. (2012). Comparative Study of Color, Pungency, and Biochemical Composition in Chili Pepper (*Capsicum annuum*) Under Different Light-emitting Diode Treatments. *HORTSCIENCE*, 47(12), 1729–1735. <https://doi.org/10.21273/HORTSCI.47.12.1729>
- García-Alcázar, M., Giménez, E., Pineda, B., Capel, C., García-Sogo, B., Sánchez, S., Yuste-Lisbona, F. J., Angosto, T., Capel, J., Moreno, V., Lozano, R. (2017). Albino T-DNA tomato mutant reveals a key function of 1-deoxy-D-xylulose-5-phosphate synthase (DXS1) in plant development and survival. *Scientific Reports*, 7, 45333. <https://doi.org/10.1038/srep45333>
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(Issue suppl_1), D258–D261. <https://doi.org/10.1093/nar/gkh036>
- Ghiasi, Z., Esmaeli, F., Aghajani, M., Ghazi-Khansari, M., Faramarzi, M. A., Amani, A. (2019). Enhancing analgesic and anti-inflammatory effects of capsaicin when loaded into olive oil nanoemulsion: An in vivo study. *International Journal of Pharmaceutics*, 559, 341–347.

<https://doi.org/10.1016/j.ijpharm.2019.01.043>

- Gong, X., Xu, Y., Li, H., Chen, X., Song, Z. (2022). Antioxidant activation, cell wall reinforcement, and reactive oxygen species regulation promote resistance to waterlogging stress in hot pepper (*Capsicum annuum* L.). *BMC Plant Biology*, 22, 425. <https://doi.org/10.1186/s12870-022-03807-2>
- González-Estrada, T., Zúñiga-Aguilar, J. J., Vázquez-Flota, F. (2018). Mejoramiento genético del chile habanero de la Península de Yucatán. CONACYT, CICY. ISBN 978-607-7823-39-1
- González-Zamora, A., Olvera-Carranza, C., Avalos-Calleros, J. A., García-Hernández, J. L., Ríos-Sánchez, E., Pérez-Morales, R. (2020). GENETIC VARIABILITY OF *Pun1* GENE (CAPSAICIN SYNTHASE) IN PUNGENT CULTIVARS OF *Capsicum annuum* OF NORTHERN MEXICO. *Tropical and Subtropical Agroecosystems*, 23(1), 26. <https://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/view/3002>
- González-Zamora, A., Sierra-Campos, E., Luna-Ortega, J. G., Pérez-Morales, R., Rodríguez-Ortiz, J. C., García-Hernández, J. L. (2013). Characterization of different *Capsicum* varieties by evaluation of their capsaicinoids content by high performance liquid chromatography, determination of pungency and effect of high temperature. *Molecules*, 18(11), 13471–13486. <https://doi.org/10.3390/molecules181113471>
- Guex, N., Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, 18(15), 2714–2723. <https://doi.org/10.1002/elps.1150181505>
- Gutiérrez-Carbajal, M. G., Monforte-González, M., de Miranda-Ham, M. L., Godoy-Hernández, G., Vázquez-Flota, F. (2010). Induction of capsaicinoid synthesis in *Capsicum chinense* cell cultures by salicylic acid or methyl jasmonate. *Biologia Plantarum*, 54(3), 430–434. <https://doi.org/10.1007/s10535-010-0078-z>

- Guzmán, I., Bosland, P. W. (2017). Sensory properties of chile pepper heat – and its importance to food quality and cultural preference. *Appetite*, 117, 186–190. <https://doi.org/10.1016/j.appet.2017.06.026>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Han, G., Qiao, Z., Li, Y., Yang, Z., Wang, C., Zhang, Y., Liu, L., Wang, B. (2022). RING Zinc Finger Proteins in Plant Abiotic Stress Tolerance. *Frontiers in Plant Science*, 13(April), 1–16. <https://doi.org/10.3389/fpls.2022.877011>
- Han, K., Jang, S., Lee, J. H., Lee, D. G., Kwon, J. K., Kang, B. C. (2019). A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. *Theoretical and Applied Genetics*, 132(4), 1235–1246. <https://doi.org/10.1007/s00122-018-03275-z>
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4, 177. [https://doi.org/https://doi.org/10.1186/1758-2946-4-17](https://doi.org/10.1186/1758-2946-4-17)
- Hepler, P. K., Winship, L. J. (2010). Calcium at the cell wall-cytoplasm interface. *Journal of Integrative Plant Biology*, 52(2), 147–160. <https://doi.org/10.1111/j.1744-7909.2010.00923.x>
- Hernández-Rodríguez, M., Rosales-Hernández, M. C., Mendieta-Wejebe, J. E., Martínez-Archundia, M., Correa Basurto, J. (2016). Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design. *Current Medicinal Chemistry*, 23(34), 3909–3924. <https://doi.org/10.2174/0929867323666160530144742>

BIBLIOGRAFÍA

- Humphrey, W., Dalke, A., Schulten, K. (1996). VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Ihsan, H., Khan, M. R., Ajmal, W., Ali, G. M. (2015). WsMAGO2, a duplicated MAGO NASHI protein with fertility attributes interacts with MPF2-like MADS-box proteins. *Planta*, 241(5), 1173–1187. <https://doi.org/10.1007/s00425-015-2247-y>
- Illumina. (2011). Quality Scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred quality scoring. https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
- Jeeatid, N., Suriharn, B., Techawongstien, S., Chanthai, S., Bosland, P. W., Techawongstien, S. (2018a). Evaluation of the effect of genotype-by-environment interaction on capsaicinoid production in hot pepper hybrids (*Capsicum chinense* Jacq.) under controlled environment. *Scientia Horticulturae*, 235, 334–339. <https://doi.org/10.1016/j.scienta.2018.03.022>
- Jeeatid, N., Techawongstien, S., Suriharn, B., Chanthai, S., Bosland, P. W., Techawongstien, S. (2018b). Influence of water stresses on capsaicinoid production in hot pepper (*Capsicum chinense* Jacq.) cultivars with different pungency levels. *Food Chemistry*, 245, 792–797. <https://doi.org/10.1016/j.foodchem.2017.11.110>
- Kabita, K. C., Sanatombi, K., Sharma, S. K. (2020). Efficient enhancement of capsaicinoids biosynthesis in cell suspension cultures of *Capsicum chinense* Jacq. cv. 'Umorok' by elicitors and differential gene expression analysis of elicited cultures. *Plant Cell, Tissue and Organ Culture*, 141, 145–154. <https://doi.org/10.1007/s11240-020-01775-9>
- Kabita, K. C., Sharma, S. K., Sanatombi, K. (2019). Analysis of capsaicinoid biosynthesis pathway genes expression in callus cultures of *Capsicum chinense* Jacq. cv. 'Umorok.' *Plant Cell, Tissue and Organ Culture*, 137, 565–573. <https://doi.org/10.1007/s11240-019-01591-w>

- Kim, J. S., Park, M., Lee, D. J., Kim, B. D. (2009). Characterization of putative capsaicin synthase promoter activity. *Molecules and Cells*, 28(4), 331–339. <https://doi.org/10.1007/s10059-009-0128-6>
- Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., Seo, E., Choi, J., Cheong, K., Kim, K. T., Jung, K., Lee, G. W., Oh, S. K., Bae, C., Kim, S. B., Lee, H. Y., Kim, S. Y., Kim, M. S., Kang, B. C., ... Choi, D. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics*, 46(3), 270–278. <https://doi.org/10.1038/ng.2877>
- Kim, T. J., Choi, J., Kim, K. W., Ahn, S. K., Ha, S. H., Choi, Y., Park, N. II, Kim, J. K. (2017). Metabolite Profiling of Peppers of Various Colors Reveals Relationships Between Tocopherol, Carotenoid, and Phytosterol Content. *Journal of Food Science*, 82(12), 2885–2893. <https://doi.org/10.1111/1750-3841.13968>
- Kirii, E., Goto, T., Yoshida, Y., Yasuba, K. I., Tanaka, Y. (2017). Non-pungency in a Japanese chili pepper landrace (*Capsicum annuum*) is caused by a novel loss-of-function *Pun1* allele. *Horticulture Journal*, 86(1), 61–69. <https://doi.org/10.2503/hortj.MI-148>
- Kirke, J., Kaplan, N., Velez, S., Jin, X. L., Vichyavichien, P., Zhang, X. H. (2018). Tissue-Preferential Activity and Induction of the Pepper Capsaicin Synthase *PUN1* Promoter by Wounding, Heat and Metabolic Pathway Precursor in Tobacco and Tomato Plants. *Molecular Biotechnology*, 60(3), 194–202. <https://doi.org/10.1007/s12033-018-0060-0>
- Koeda, S., Sato, K., Saito, H., Nagano, A. J., Yasugi, M., Kudoh, H., Tanaka, Y. (2019). Mutation in the putative ketoacyl-ACP reductase *CaKR1* induces loss of pungency in *Capsicum*. *Theoretical and Applied Genetics*, 132(1), 65–80. <https://doi.org/10.1007/s00122-018-3195-2>
- Koeda, S., Sato, K., Tomi, K., Tanaka, Y., Takisawa, R., Hosokawa, M., Doi, M., Nakazaki, T., Kitajima, A. (2014). Analysis of non-pungency, aroma, and origin of a *Capsicum chinense*

BIBLIOGRAFÍA

cultivar from a Caribbean Island. *Journal of the Japanese Society for Horticultural Science*, 83(3), 244–251. <https://doi.org/10.2503/jjshs1.CH-105>

Krebs, J. E., Goldstein, E. S., Kilpatrick, S. T. (2018). *Lewin's Genes XII* (12th ed.). Jones & Bartlett Learning.

Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications, en *Next Generation Sequencing - Advances, Applications and Challenges*, Kulski, J. (Ed.). InTech. <https://doi.org/10.5772/61964>

Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>

Laflamme, P., St.-Pierre, B., De Luca, V. (2001). Molecular and biochemical analysis of a Madagascar periwinkle root-specific minovincinine-19-hydroxy-o-acetyltransferase. *Plant Physiology*, 125(1), 189–198. <https://doi.org/10.1104/pp.125.1.189>

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>

Latournerie-Moreno, L., Lopez-Vázquez, J. ., Castañón-Nájera, G., Mijangos-Cortes, J. O., Espadas-Villamil, G., Pérez-Gutiérrez, A., Ruiz-Sánchez, E. (2015). AGRONOMIC EVALUATION OF HABANERO PEPPER (*Capsicum chinense* Jacq.). *Agroproductividad*, 8(1), 24–29.

Lemos, V. C., Reimer, J. J., Wormit, A. (2019). Color for life: Biosynthesis and distribution of phenolic compounds in pepper (*Capsicum annuum*). *Agriculture*, 9(4), 81. <https://doi.org/10.3390/agriculture9040081>

- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., Rouzé, P., Rombauts, S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research*, 30(1), 325–327. <https://doi.org/10.1093/nar/30.1.325>
- Leszczuk, A., Kalaitzis, P., Blazakis, K. N., Zdunek, A. (2020). The role of arabinogalactan proteins (AGPs) in fruit ripening—a review. *Horticulture Research*, 7, 176. <https://doi.org/10.1038/s41438-020-00397-8>
- Leung, E. W. W., Guddat, L. W. (2009). Conformational Changes in a Plant Ketol-Acid Reductoisomerase upon Mg²⁺ and NADPH Binding as Revealed by Two Crystal Structures. *Journal of Molecular Biology*, 389(1), 167–182. <https://doi.org/10.1016/j.jmb.2009.04.012>
- Li, B., Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Lin, R. J., Wu, I. J., Hong, J. Y., Liu, B. H., Liang, R. Y., Yuan, T. M., Chuang, S. M. (2018). Capsaicin-induced TRIB3 upregulation promotes apoptosis in cancer cells. *Cancer Management and Research*, 10, 4237–4248. <https://doi.org/10.2147/CMAR.S162383>
- Liu, S., Chen, C., Chen, G., Cao, B., Chen, Q., Lei, J. (2012). RNA-sequencing tag profiling of the placenta and pericarp of pungent pepper provides robust candidates contributing to capsaicinoid biosynthesis. *Plant Cell, Tissue and Organ Culture*, 110(1), 111–121. <https://doi.org/10.1007/s11240-012-0135-8>
- Liu, S., Li, W., Wu, Y., Chen, C., Lei, J. (2013). *De Novo* Transcriptome Assembly in Chili Pepper (*Capsicum frutescens*) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. *PLoS ONE*, 8(1), e48156. <https://doi.org/10.1371/journal.pone.0048156>

BIBLIOGRAFÍA

- Lokstein, H., Renger, G., Götze, J. P. (2021). Photosynthetic light-harvesting (antenna) complexes-structures and functions. *Molecules*, 26(11), 3378. <https://doi.org/10.3390/molecules26113378>
- Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Luo, W., Brouwer, C. (2013). Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14), 1830–1831. <https://doi.org/10.1093/bioinformatics/btt285>
- Lupi, A. C. D., Lira, B. S., Gramegna, G., Trench, B., Alves, F. R. R., Demarco, D., Peres, L. E. P., Purgatto, E., Freschi, L., Rossi, M. (2019). *Solanum lycopersicum* GOLDEN 2-LIKE 2 transcription factor affects fruit quality in a light- and auxin-dependent manner. *PLoS ONE*, 14(2), e0212224. <https://doi.org/10.1371/journal.pone.0212224>
- Ma, X., Koepke, J., Panjikar, S., Fritsch, G., Stöckigt, J. (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *Journal of Biological Chemistry*, 280(14), 13576–13583. <https://doi.org/10.1074/jbc.M414508200>
- Ma, X., Qiao, Y. M., Li, Y., Yu, Y. N., Gong, Z. H. (2022). Identification of Fruit Traits Related QTLs and a Candidate Gene, *CaBRX*, Controlling Locule Number in Pepper (*Capsicum annuum* L.). *Horticulturae*, 8(2), 146. <https://doi.org/10.3390/horticulturae8020146>
- Manjasetty, B. A., Yu, X. H., Panjikar, S., Taguchi, G., Chance, M. R., Liu, C. J. (2012). Structural basis for modification of flavonol and naphthol glucoconjugates by *Nicotiana tabacum* malonyltransferase (*NtMaT1*). *Planta*, 236(3), 781–793. <https://doi.org/10.1007/s00425-012-1660-8>

- Martin, J. A., Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12(10), 671–682. <https://doi.org/10.1038/nrg3068>
- Martínez-López, L. A., Ochoa-Alejo, N., Martínez, O. (2014). Dynamics of the chili pepper transcriptome during fruit development. *BMC Genomics*, 15, 143. <https://doi.org/10.1186/1471-2164-15-143>
- McDermaid, A., Monier, B., Zhao, J., Liu, B., Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics*, 20(6), 2044–2054. <https://doi.org/10.1093/bib/bby067>
- Medina-Lara, F., Echevarría-Machado, I., Pacheco-Arjona, R., Ruiz-Lau, N., Guzmán-Antonio, A., & Martínez-Estévez, M. (2008). Influence of Nitrogen and Potassium Fertilization on Fruiting and Capsaicin Content in Habanero Pepper (*Capsicum chinense* Jacq.). *HORTSCIENCE*, 43(5), 1549–1554. <https://doi.org/10.21273/HORTSCI.43.5.1549>
- Medina-Lara, F., Souza-Perera, R., Martínez-Estévez, M., Ramírez-Sucre, M. O., Rodríguez-Buenfil, I. M., Machado, I. E. (2019). Red and brown soils increase the development and content of nutrients in habanero pepper subjected to irrigation water with high electrical conductivity. *HortScience*, 54(11), 2039–2049. <https://doi.org/10.21273/HORTSCI14157-19>
- Mokhtar, M., Ginestra, G., Youcefi, F., Filocamo, A., Bisignano, C., Riazi, A. (2017). Antimicrobial Activity of Selected Polyphenols and Capsaicinoids Identified in Pepper (*Capsicum annuum* L.) and Their Possible Mode of Interaction. *Current Microbiology*, 74(11), 1253–1260. <https://doi.org/10.1007/s00284-017-1310-2>
- Monforte-González, M., Guzmán-Antonio, A., Uuh-Chim, F., Vázquez-Flota, F. (2010). Capsaicin accumulation is related to nitrate content in placentas of habanero peppers (*Capsicum chinense* Jacq.). *Journal of the Science of Food and Agriculture*, 90(5), 764–768. <https://doi.org/10.1002/jsfa.3880>

BIBLIOGRAFÍA

- Morales-Quintana, L., Fuentes, L., Gaete-Eastman, C., Herrera, R., Moya-León, M. A. (2011). Structural characterization and substrate specificity of VpAAT1 protein related to ester biosynthesis in mountain papaya fruit. *Journal of Molecular Graphics and Modelling*, 29(5), 635–642. <https://doi.org/10.1016/j.jmglm.2010.11.011>
- Morales-Quintana, L., Moya-León, M. A., Herrera, R. (2012). Molecular docking simulation analysis of alcohol acyltransferases from two related fruit species explains their different substrate selectivities. *Molecular Simulation*, 38(11), 912–921. <https://doi.org/10.1080/08927022.2012.672738>
- Naves, E. R., de Ávila Silva, L., Sulpice, R., Araújo, W. L., Nunes-Nesi, A., Peres, L. E. P., Zsögön, A. (2019). Capsaicinoids: Pungency beyond *Capsicum*. *Trends in Plant Science*, 24(2), 109–120. <https://doi.org/10.1016/j.tplants.2018.11.001>
- Nelson, D. L., Cox, M. M. (2014). Lehninger. Principios de bioquímica (6th ed.). Omega.
- Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palette. <https://cran.r-project.org/package=RColorBrewer>
- Nguyen, T. M., Shafi, A., Nguyen, T., Draghici, S. (2019). Identifying significantly impacted pathways: A comprehensive review and assessment. *Genome Biology*, 20, 203. <https://doi.org/10.1186/s13059-019-1790-4>
- Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., Giovannoni, J. J. (2014). Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, 26(2), 585–601. <https://doi.org/10.1105/tpc.113.118794>
- Nieto-Garibay, A., Barraza, A., Caamal-Chan, G., Murillo-Amador, B., Troyo-Diéiguez, E., Burgoa-Cruz, C. A., Jaramillo-Limón, J. N., Loera-Muro, A. (2022). Habanero pepper (*Capsicum*

chinense) adaptation to water-deficit stress in a protected agricultural system. *Functional Plant Biology*, 49(3), 295–306. <https://doi.org/10.1071/FP20394>

Ogawa, K., Murota, K., Shimura, H., Furuya, M., Togawa, Y., Matsumura, T., Masuta, C. (2015). Evidence of capsaicin synthase activity of the *Pun1*-encoded protein and its role as a determinant of capsaicinoid accumulation in pepper. *BMC Plant Biology*, 15(1), 93. <https://doi.org/10.1186/s12870-015-0476-7>

Olguín-Rojas, J. A., Fayos, O., Vázquez-León, L. A., Ferreiro-González, M., del Carmen Rodríguez-Jimenes, G., Palma, M., Garcés-Claver, A., Barbero, G. F. (2019). Progression of the total and individual capsaicinoids content in the fruits of three different cultivars of *Capsicum chinense* Jacq. *Agronomy*, 9(3), 141. <https://doi.org/10.3390/agronomy9030141>

Orellana-Lanza, R., Espadas-Manrique, C., Nava-Marín, F. (2010). Capítulo 1. Contexto físico. Climas, en: Biodiversidad y desarrollo humano en Yucatán, Durán, R, Méndez, M. (Eds.). CICY, PPD-FMAM, CONABIO, SEDUMA.

Palma, J. M., Terán, F., Contreras-Ruiz, A., Rodríguez-Ruiz, M., Corpas, F. J. (2020). Antioxidant profile of pepper (*Capsicum annuum* L.) fruits containing diverse levels of capsaicinoids. *Antioxidants*, 9(9), 878. <https://doi.org/10.3390/antiox9090878>

Park, J., Cui, Y., Kang, B. H. (2015b). AtPGL3 is an Arabidopsis BURP domain protein that is localized to the cell wall and promotes cell enlargement. *Frontiers in Plant Science*, 6, 412. <https://doi.org/10.3389/fpls.2015.00412>

Park, M., Lee, J. H., Han, K., Jang, S., Han, J., Lim, J. H., Jung, J. W., Kang, B. C. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of *Capsicum chinense* revealed using QTL-seq and RNA-seq. *Theoretical and Applied Genetics*, 132(2), 515–529. <https://doi.org/10.1007/s00122-018-3238-8>

BIBLIOGRAFÍA

- Park, Y.-J., Nishikawa, T., Minami, M., Nemoto, K., Iwasaki, T., Matsushima, K. (2015a). A low-pungency S3212 genotype of *Capsicum frutescens* caused by a mutation in the putative aminotransferase (*p-AMT*) gene. *Molecular Genetics and Genomics*, 290(6), 2217–2224. <https://doi.org/10.1007/s00438-015-1071-1>
- Peña-Yam, L. P., Muñoz-Ramírez, L. S., Avilés-Viñas, S. A., Canto-Flick, A., Pérez-Pastrana, J., Guzmán-Antonio, A., Santana-Buzzy, N., Aguilera-Cauich, E. A., Mijangos-Cortés, J. O. (2019). Analysis of genetic parameters of habanero pepper (*Capsicum chinense* jacq.) in the Yucatan, Mexico. *HortScience*, 54(3), 429–433. <https://doi.org/10.21273/HORTSCI13710-18>
- Phimchan, P., Chanthai, S., Bosland, P. W., Techawongstien, S. (2014). Enzymatic changes in phenylalanine ammonia-lyase, cinnamic-4-hydroxylase, capsaicin synthase, and peroxidase activities in *Capsicum* under drought stress. *Journal of Agricultural and Food Chemistry*, 62(29), 7057–7062. <https://doi.org/10.1021/jf4051717>
- Powell, A. L. T., Nguyen, C. v., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernandes-Muñoz, R., Vicente, A., Lopez-Baltazar, J., Barry, C. S., Liu, Y., Chetelat, R., Granell, A., Deynze, A. Van, Giovannoni, J. J., Bennett, A. B. (2012). Uniform ripening Encodes a *Golden 2-like* Transcription Factor Regulating Tomato Fruit Chloroplast Development. *Science*, 336, 1711–1715. <https://doi.org/10.1126/science.1222218>
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Changying, L. H., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D. L., ... Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14), 5135–5140. <https://doi.org/10.1073/pnas.1400975111>
- Raghavan, V., Kraft, L., Mesny, F., Rigerte, L. (2022). A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics*, 23(2), bbab563.

<https://doi.org/10.1093/bib/bbab563>

- Ramírez Meraz, M., Arcos Cavazos, G., Méndez Aguilar, R. (2018). Jaguar: cultivar de chile habanero para México. *Revista Mexicana de Ciencias Agrícolas*, 9(2), 487–492. <https://doi.org/10.29312/remexca.v9i2.1089>
- Reddy, U. K., Almeida, A., Abburi, V. L., Alaparathi, S. B., Unselt, D., Hankins, G., Park, M., Choi, D., Nimmakayala, P. (2014). Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in *Capsicum annuum* L. collections. *PLoS ONE*, 9(1), 1–10. <https://doi.org/10.1371/journal.pone.0086393>
- Ribeiro, S. F. F., Taveira, G. B., Carvalho, A. O., Dias, G. B., Cunha, M. Da, Santa-Catarina, C., Rodrigues, R., Gomes, V. M. (2012). Antifungal and other biological activities of two 2S albumin-homologous proteins against pathogenic fungi. *Protein Journal*, 31(1), 59–67. <https://doi.org/10.1007/s10930-011-9375-4>
- Rodas-Junco, B. A., Cab-Guillén, Y., Muñoz-Sánchez, J. A., Vázquez-Flota, F., Monforte-González, M., Hernández-Sotomayor, S. M. T. (2013). Salicylic acid induces vanillin synthesis through the phospholipid signaling pathway in *Capsicum chinense* cell cultures. *Plant Signaling and Behavior*, 8(10), e26752. <https://doi.org/10.4161/psb.26752>
- Rödiger, A., Agne, B., Dobritsch, D., Helm, S., Müller, F., Pöttsch, N., Baginsky, S. (2021). Chromoplast differentiation in bell pepper (*Capsicum annuum*) fruits. *Plant Journal*, 105(5), 1431–1442. <https://doi.org/10.1111/tpj.15104>
- Rodríguez-Buenfil, I., López-Domínguez, C., Morozova, K., Schampicchio, M., Ramírez, Sucre, M. (2018). Capsaicinoids and vitamin C in habanero chili (*Capsicum chinense*) cultivated in different types of soils from Yucatan, Mexico. *2nd EuroSciCon Conference on Food Technology*, 2, 36–36. <https://doi.org/10.21767/2577-0586-c1-002>

BIBLIOGRAFÍA

- Roy, A., Kucukural, A., Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Ruiz-Lau, N., Medina-Lara, F., Martínez-Estévez, M. (2011). El chile habanero: su origen y usos. *Ciencia-Academia Mexicana de Ciencias*, 62(3), 70–78. http://www.revistaciencia.amc.edu.mx/images/revista/62_3/PDF/Habanero.pdf
- Schaetzl, R.-J., Anderson, S. (2005). Basics concepts: soil morphology, en *Soils. Genesis and Geomorphology*. Cambridge University Press, pp. 9-31.
- Schöning-Stierand, K., Diedrich, K., Fährrolfes, R., Flachsenberg, F., Meyder, A., Nittinger, E., Steinegger, R., Rarey, M. (2020). ProteinsPlus: Interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Research*, 48(W1), W48–W53. <https://doi.org/10.1093/NAR/GKAA235>
- Silva, J., Kim, Y. J., Sukweenadhi, J., Rahimi, S., Kwon, W. S., Yang, D. C. (2016). Molecular characterization of 5-chlorophyll a/b-binding protein genes from *Panax ginseng* Meyer and their expression analysis during abiotic stresses. *Photosynthetica*, 54(3), 446–458. <https://doi.org/10.1007/s11099-016-0189-7>
- Soares, L. M. M., Valcárcel, J. (2006). The expanding transcriptome: The genome as the “Book of Sand.” *EMBO Journal*, 25(5), 923–931. <https://doi.org/10.1038/sj.emboj.7601023>
- Sommer-Knudsen, J., Clarke, A. E., Bacic, A. (1996). A galactose-rich, cell-wall glycoprotein from styles of *Nicotiana glauca*. *The Plant Journal*, 9(1), 71–83. <https://doi.org/10.1046/j.1365-3113.1996.09010071.x>
- Song, Y., Pu, H., Jiang, T., Zhang, L., Ouyang, M. (2016). Crystal structure of glutamate-1-semialdehyde-2,1-aminomutase from *Arabidopsis thaliana*. *Acta Crystallographica*

Section:F Structural Biology Communications, 72, 448–456.
<https://doi.org/10.1107/S2053230X16007263>

Sosa-Moguel, O., Pino, J. A., Ayora-Talavera, G., Sauri-Duch, E., Cuevas-Glory, L. (2018). Biological activities of volatile extracts from two varieties of Habanero pepper (*Capsicum chinense* Jacq.). *International Journal of Food Properties*, 20(S3), S3042–S3051. <https://doi.org/10.1080/10942912.2017.1397694>

Srinivasan, K. (2016). Biological Activities of Red Pepper (*Capsicum annuum*) and Its Pungent Principle Capsaicin: A Review. *Critical Reviews in Food Science and Nutrition*, 56(9), 1488–1500. <https://doi.org/10.1080/10408398.2013.772090>

Stellari, G. M., Mazourek, M., Jahn, M. M. (2010). Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity*, 104(5), 460–471. <https://doi.org/10.1038/hdy.2009.131>

Stewart Jr, C., Kang, B. C., Liu, K., Mazourek, M., Moore, S. L., Eun, Y. Y., Kim, B. D., Paran, I., Jahn, M. M. (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal*, 42(5), 675–688. <https://doi.org/10.1111/j.1365-313X.2005.02410.x>

Stewart Jr, C., Mazourek, M., Stellari, G. M., O'Connell, M., Jahn, M. (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany*, 58(5), 979–991. <https://doi.org/10.1093/jxb/erl243>

Sugiyama, R. (2017). Capsaicinoids production and accumulation in epidermal cells on the internal side of the fruit pericarp in 'Bhut Jolokia' (*Capsicum chinense*). *Cytologia*, 82(3), 303–306. <https://doi.org/10.1508/cytologia.82.303>

Sun, B., Chen, C., Song, J., Zheng, P., Wang, J., Wei, J., Cai, W., Chen, S., Cai, Y., Yuan, Y.,

BIBLIOGRAFÍA

- Zhang, S., Liu, S., Lei, J., Cheng, G., Zhu, Z. (2022). The *Capsicum* MYB31 regulates capsaicinoid biosynthesis in the pepper pericarp. *Plant Physiology and Biochemistry*, 176, 21–30. <https://doi.org/10.1016/j.plaphy.2022.02.014>
- Sun, B., Zhou, X., Chen, C., Chen, C., Chen, K., Chen, M., Liu, S., Chen, G., Cao, B., Cao, F., Lei, J., Zhu, Z. (2020). Coexpression network analysis reveals an MYB transcriptional activator involved in capsaicinoid biosynthesis in hot peppers. *Horticulture Research*, 7, 162. <https://doi.org/10.1038/s41438-020-00381-2>
- Sun, B., Zhu, Z., Chen, C., Chen, G., Cao, B., Chen, C., Lei, J. (2019). Jasmonate-Inducible R2R3-MYB Transcription Factor Regulates Capsaicinoid Biosynthesis and Stamen Development in *Capsicum*. *Journal of Agricultural and Food Chemistry*, 67(39), 10891–10903. <https://doi.org/10.1021/acs.jafc.9b04978>
- Suzuki, T., Kawada, T., Iwai, K. (1981). Biosynthesis of acyl moieties of capsaicin and its analogues from valine and leucine in *Capsicum* fruits. *Plant and Cell Physiology*, 22(1), 23–32. <https://doi.org/10.1093/oxfordjournals.pcp.a076142>
- Tanaka, Y., Asano, T., Kanemitsu, Y., Goto, T., Yoshida, Y., Yasuba, K., Misawa, Y., Nakatani, S., Kobata, K. (2019). Positional differences of intronic transposons in *pAMT* affect the pungency level in chili pepper through altered splicing efficiency. *Plant Journal*, 100(4), 693–705. <https://doi.org/10.1111/tpj.14462>
- Tanaka, Y., Fukuta, S., Koeda, S., Goto, T., Yoshida, Y., Yasuba, K. (2018). Identification of a novel mutant *pAMT* allele responsible for low-pungency and capsinoid production in chili pepper: Accession 'no. 4034' (*Capsicum chinense*). *Horticulture Journal*, 87(2), 222–228. <https://doi.org/10.2503/hortj.OKD-115>
- Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T., Yazawa, S. (2010). Novel loss-of-function putative aminotransferase alleles cause biosynthesis of capsinoids, nonpungent capsaicinoid analogues, in mildly pungent chili peppers (*Capsicum chinense*). *Journal of*

Agricultural and Food Chemistry, 58(22), 11762–11767. <https://doi.org/10.1021/jf1019642>

Tanaka, Y., Nakashima, F., Kirii, E., Goto, T., Yoshida, Y., Yasuba, K.-ichiro. (2017). Difference in capsaicinoid biosynthesis gene expression in the pericarp reveals elevation of capsaicinoid contents in chili peppers (*Capsicum chinense*). *Plant Cell Reports*, 36(2), 267–279. <https://doi.org/10.1007/s00299-016-2078-8>

Tanaka, Y., Sonoyama, T., Muraga, Y., Koeda, S., Goto, T., Yoshida, Y., Yasuba, K. (2015). Multiple loss-of-function putative aminotransferase alleles contribute to low pungency and capsinoid biosynthesis in *Capsicum chinense*. *Molecular Breeding*, 35, 142. <https://doi.org/10.1007/s11032-015-0339-9>

Tanaka, Y., Watachi, M., Nemoto, W., Goto, T., Yoshida, Y., Yasuba, K.-ichiro, Ohno, S., Doi, M. (2021). Capsaicinoid biosynthesis in the pericarp of chili pepper fruits is associated with a placental septum-like transcriptome profile and tissue structure. *Plant Cell Reports*, 40(10), 1859–1874. <https://doi.org/10.1007/s00299-021-02750-0>

Thiele, R., Mueller-Seitz, E., Petz, M. (2008). Chili pepper fruits: Presumed precursors of fatty acids characteristic for capsaicinoids. *Journal of Agricultural and Food Chemistry*, 56(11), 4219–4224. <https://doi.org/10.1021/jf073420h>

Tripodi, P., Kumar, S. (2019). The *Capsicum* Crop: An Introduction, en: The *Capsicum* Genome. Compendium of Plant Genomes, Ramchiary, N., Kole, C. (Eds.). Springer, Cham. https://doi.org/10.1007/978-3-319-97217-6_1

Trott, O., Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461. <https://doi.org/10.1002/jcc.21334>

Tsurumaki, K., Sasanuma, T. (2019). Discovery of novel unfunctional *pAMT* allele *pamt^{t10}* causing

BIBLIOGRAFÍA

loss of pungency in sweet bell pepper (*Capsicum annuum* L.). *Breeding Science*, 69(1), 133–142. <https://doi.org/10.1270/jsbbs.18150>

Tyler, B. J., Peterson, R. E., Lee, T. G., Draude, F., Pelster, A., Arlinghaus, H. F. (2016). ToF-SIMS imaging of capsaicinoids in Scotch Bonnet peppers (*Capsicum chinense*). *Biointerphases*, 11(2), 02A327. <https://doi.org/10.1116/1.4945326>

Vaishnavi, B. A., Bhoomika, H. R., Raviraj Shetty, G. (2018). Genetic Parameters Study for Growth, Yield and Quality Traits in Bird's Eye Chilli (*Capsicum frutescens* L.). *International Journal of Current Microbiology and Applied Sciences*, 7(5), 1813–1817. <https://doi.org/10.20546/ijcmas.2018.705.213>

Villa Castorena, M., Catalán-Valencia, A. E., Inzunza-Ibarra, M. A., Román López, A., Macías Rodríguez, H., Cabrera Rodarte, D. (2014). PRODUCCIÓN HIDROPÓNICA DE CHILE HABANERO EN INVERNADERO (1st ed). INIFAP, SAGARPA. <http://www.inifap.gob.mx>

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. *Journal of Chemical Information and Modeling*, 52(2), 360–372. <https://doi.org/10.1021/ci200454v>

Walker, A. M., Hayes, R. P., Youn, B., Vermerris, W., Sattler, S. E., Kang, C. (2013). Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme a-dependent transferases and synthases. *Plant Physiology*, 162(2), 640–651. <https://doi.org/10.1104/pp.113.217836>

Wang, F., Xue, Y., Fu, L., Wang, Y., He, M., Zhao, L., Liao, X. (2022). Extraction, purification, bioactivity and pharmacological effects of capsaicin: a review. *Critical Reviews in Food Science and Nutrition*, 62(19), 5322–5348. <https://doi.org/10.1080/10408398.2021.1884840>

- Wang, J., Liu, Y., Tang, B., Dai, X., Xie, L., Liu, F., Zou, X. (2020). Genome-Wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the R2R3-MYB Gene Family in *Capsicum annuum* L. *Frontiers in Genetics*, 11, 598183. <https://doi.org/10.3389/fgene.2020.598183>
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. (2022). gplots: Various R Programming Tools for Plotting Data. <https://cran.r-project.org/package=gplots>
- Weber, N., Ismail, A., Gorwa-Grauslund, M., Carlquist, M. (2014). Biocatalytic potential of vanillin aminotransferase from *Capsicum chinense*. *BMC Biotechnology*, 14, 25. <https://doi.org/10.1186/1472-6750-14-25>
- Werner, J. (2021). Capsaicinoids – Properties and Mechanisms of Pro-health Action, en: Analytical Methods in the Determination of Bioactive Compounds and Elements in Food. Food Bioactive Ingredients, Jeszka-Skowron, M., Zgola-Grześkowiak, A., Grześkowiak, T., Ramakrishna, A. (Eds.). Springer Nature, pp. 193-225. <https://doi.org/10.1007/978-3-030-61879-7>
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Xu, R., Li, Q. Q. (2003). A RING-H2 zinc-finger protein gene *RIE1* is essential for seed development in *Arabidopsis*. *Plant Molecular Biology*, 53, 37–50. <https://doi.org/10.1023/B:PLAN.0000009256.01620.a6>

BIBLIOGRAFÍA

- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, 12(1), 7–8. <https://doi.org/10.1038/nmeth.3213>
- Yang, J., Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Current Protocols in Bioinformatics*, 52(1), 5.8.1-5.8.15. <https://doi.org/10.1002/0471250953.bi0508s52>
- Yu, G. (2022). enrichplot: Visualization of Functional Enrichment Result. R package version 1.16.2. <https://yulab-smu.top/biomedical-knowledge-mining-book/>
- Zamski, E., Shoham, O., Palevitch, D., Levy, A. (1987). Ultrastructure of capsaicinoid-secreting cells in pungent and nonpungent red pepper (*Capsicum annuum* L.) cultivars. *Botanical Gazette*, 148(1), 1–6. <https://doi.org/10.1086/337620>
- Zamudio-Moreno, E., Echevarría-Machado, I., Medina-Lara, M. de F., Calva-Calva, G., Miranda-Ham, M. de L., Martínez-Estévez, M. (2014). Role of peroxidases in capsaicinoids degradation in habanero pepper (*Capsicum chinense* Jacq.) plants grown under water deficit conditions. *AJCS*, 8(3), 448–454.
- Zhang, J., Lv, J., Xie, J., Gan, Y., Coulter, J. A., Yu, J., Li, J., Wang, J., Zhang, X. (2020a). Nitrogen Source Affects the Composition of Metabolites in Pepper (*Capsicum annuum* L.) and Regulates the Synthesis of Capsaicinoids through the GOGAT-GS Pathway. *Foods*, 9(2), 150. <https://doi.org/10.3390/foods9020150>
- Zhang, S., Ding, G., He, W., Liu, K., Luo, Y., Tang, J., He, N. (2020b). Functional Characterization of the 1-Deoxy-D-Xylulose 5-Phosphate Synthase Genes in *Morus notabilis*. *Frontiers in Plant Science*, 11, 1142. <https://doi.org/10.3389/fpls.2020.01142>
- Zhang, X., Gonzalez-Carranza, Z. H., Zhang, S., Miao, Y., Liu, C. J., Roberts, J. A. (2019). F-box

proteins in plants. *Annual Plant Reviews Online*, 2(1), 307–327.
<https://doi.org/10.1002/9781119312994.apr0701>

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40.
<https://doi.org/10.1186/1471-2105-9-40>

Zhang, Z.-X., Zhao, S.-N., Liu, G.-F., Huang, Z.-M., Cao, Z.-M., Cheng, S.-H., Lin, S.-S. (2016). Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. *Scientific Reports*, 6, 34121.
<https://doi.org/10.1038/srep34121>

Zhao, S., Gao, H., Luo, J., Wang, H., Dong, Q., Wang, Y., Yang, K., Mao, K., Ma, F. (2020). Genome-wide analysis of the light-harvesting chlorophyll a/b-binding gene family in apple (*Malus domestica*) and functional characterization of *MdLhcb4.3*, which confers tolerance to drought and osmotic stress. *Plant Physiology and Biochemistry*, 154, 517–529.
<https://doi.org/10.1016/j.plaphy.2020.06.022>

Zheng, W., Zhang, C., Bell, E. W., Zhang, Y. (2019). I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. *Future Generation Computer Systems*, 99, 73–85. <https://doi.org/10.1016/j.future.2019.04.011>

Zhu, Z., Sun, B., Cai, W., Zhou, X., Mao, Y., Chen, C., Wei, J., Cao, B., Chen, C., Chen, G., Lei, J. (2019). Natural variations in the MYB transcription factor *MYB31* determine the evolution of extremely pungent peppers. *New Phytologist*, 223(2), 922–938.
<https://doi.org/10.1111/nph.15853>

ANEXOS

ANEXOS

TABLAS SUPLEMENTARIAS

Tabla suplementaria 3.1. Información de genes diferencialmente expresados en frutos de 25 DPA de plantas de chile habanero crecidas en suelo rojo o suelo negro. Se presenta el nombre del *contig* y su clave de acuerdo a los unigenes seleccionados en la gráfica de volcán Figura 3.5A. A) Genes más expresados en suelo rojo ($p\text{value} < 0.00835$ y $\text{LFC} \geq 1.5$), y B) genes más expresados en suelo negro ($p\text{value} < 0.00835$ y $\text{LFC} \leq -1.5$)

A. Genes más expresados en suelo rojo

* <i>Contig</i> (clave)	Longitud(nt)	Descripción	Anotación ID	Bit score	Cober tura q. (%)	Cober tura t. (%)	Identi dad (%)	Valor e	SN1 cuen tas	SN2 cuen tas	SR1 cuen tas	SR2 cuen tas	LFC	pval ue	padj
c65010_ g13_i1 (1)	1864	No anotado							1	1	67	41	5.55	7.95 E-06	8.58E -04
c67378_ g1_i1 (2)	405	3-cetoacil- CoA sintasa	XM_0167 05237	708.4	96.05	18.75	99	0	3	0	119	43	5.51	2.33 E-07	5.43 E-05
c62416_ g3_i1 (3)	341	Proteína F- box	cann:1078 71265	67.8	27.57	18.9	100	2 E-09	12	3	302	322	5.16	2.65 E-23	6.07E -19
c63084_ g2_i1 (4)	501	Calmodulin a lisina N- metiltransfe rasa	nta:10779 5679	68.0	25.34	10.09	72	7 E-08	1	2	55	46	4.88	8.59 E-06	9.11E -04
c36776_ g1_i1 (5)	351	Proteoglu cano rico en prolina	XM_0167 07643	501.6	79.77	15.45	98	5E- 138	1	3	65	40	4.53	8.64 E-06	9.11E -04

c66905_g1_i1 (6)	638	Extensina específica de pistilo	PHT9043 4.1	96.3	37.3	28.78	93	1E-16	126	246	131 2	990	2.46	1.78 E-19	2.04E -15
c80250_g1_i1 (7)	449	Poliubiquitina	XM_0166 87876	446.2	56.57	13.4	98	3E-121	388	588	214 5	168 8	1.79	4.75 E-16	2.18E -12
c109463_g1_i1 (8)	551	Dominio de proteína BURP	cann:1078 70425	270.8	98.73	49.32	97	3E-70	48	74	410	352	2.46	2.36 E-14	6.75E -11
c58770_g1_i1 (9)	4029	Proteína cinasa tipo receptor de repetición rico en leucina	PHT2591 9.1	216.9	20.35	28.93	49	4E-52	511	682	213 1	176 7	1.52	1.24 E-13	3.16E -10
c63235_g1_i1 (10)	3367	Proteína F-box	sstn:1258 56603	735	18.59	50.97	86	0	205	279	143 8	798	2.01	5.03 E-13	1.15E -09
c65909_g6_i1 (11)	608	Extensina	PHT8514 5.1	226.9	69.24	12.24	84	6E-56	248	316	956	108 4	1.67	6.97 E-13	1.33E -09
c63526_g1_i1 (12)	271	Proteína mago nashi (homólogo)	XM_0167 03319	444.3	89.67	30.38	99	7E-121	86	113	512	420	2.03	1.55 E-12	2.41E -09

ANEXOS

c24022_g2_i1 (13)	596	Proteína hipotética	EPS7450 5.1	253.1	64.09	50.2	99	7E-64	480	891	4156	2132	2.02	1.58E-12	2.41E-09
-------------------	-----	---------------------	-------------	-------	-------	------	----	-------	-----	-----	------	------	------	----------	----------

B. Genes más expresados en suelo negro

*Contig (clave)	Longitud(nt)	Descripción	Anotación ID	Bit score	Coertura q. (%)	Coertura t. (%)	Identidad (%)	Valor e	SN1 cue ntas	SN2 cue ntas	SR1 cue ntas	SR2 cue ntas	LFC	pvalue	padj
c107830_g1_i1 (14)	2027	Similar a albúmina	NP_03378 4.2	1256.5	90.03	100	100	0	1	916	2	1	-8.33	2.58E-03	4.87E-02
c61203_g2_i1 (15)	1092	Proteína de unión 3C a clorofila	cann:1078 57025	549.7	73.44	100	100	6E-154	23683	28996	6149	9339	-1.95	6.53E-18	4.99E-14
c54446_g1_i1 (16)	1297	Proteína de dedo de anillo H2	KAF3637 013.1	457	56.90	93.18	97.15	3E-158	611	867	213	210	-1.99	3.10E-16	1.77E-12
c65505_g9_i1 (17)	1108	BREVIS RADIX	cann:1078 69944	360.1	46.39	50.74	100	7E-97	1086	1624	446	478	-1.73	6.59E-15	2.52E-11
c50314_g1_i1 (18)	1245	Proteína de unión 6A a clorofila	PHT4920 2.1	512.3	59.36	100	100	1E-141	7943	10054	3160	3938	-1.53	1.13E-14	3.70E-11
c66518_g5_i2 (19)	905	Proteína de unión 1B a clorofila	cann:1078 55767	493.4	80.33	91.32	100	4E-137	6305	8017	1834	3004	-1.74	6.96E-13	1.33E-09

c65207_g6_i1 (20)	7996	Proteína hipotética mitocondria I. Dominio de transcriptasa inversa	cann:1998 9150	311.2	5.72	100	100	3E-81	105 1	216 2	487	591	- 1.73	1.17 E-10	8.65E -08
c66396_g5_i2 (21)	2430	Proteína hipotética	PHT5237 0.1	233.4	18.68	72.07	80	3E-57	954	166 9	432	544	- 1.59	3.16 E-10	2.13E -07
c65950_g13_i1 (22)	1828	Proteína hipotética mitocondria I	cann:1998 9070	442.2	39.77	100	100	2E-121	379	735	156	190	- 1.85	1.00 E-09	5.73E -07
c66138_g3_i1 (23)	1659	Proteína hipotética mitocondria I	cann:1998 9083	240.4	21.22	100	100	1E-60	334	835	157	180	- 1.94	7.38 E-09	2.91E -06
c109544_g1_i1 (24)	552	Ribulosa bifosfato carboxilasa /oxigenasa activasa (RuBisCO)	PHU0537 1.1	309.3	84.42	35.15	100	8E-81	398	340	81	137	- 1.96	1.17 E-08	4.25E -06

*La identificación de la anotación corresponde a la base de datos de KEGG o del NCBI, de acuerdo a la mejor anotación que tuvo el *contig*. La cobertura q. corresponda a la cobertura del *contig* ("query"). La cobertura t. corresponde a la cobertura de la secuencia de referencia de bases de datos "target"

ANEXOS

Tabla Suplementaria 3.2. Anotación funcional de genes seleccionados de las rutas metabólicas enriquecidas y de la síntesis de capsaicinoides, y datos de expresión de los genes en TPM (transcritos por millón) para realizar el mapa de calor. Los genes están ordenados por sus TPM y de acuerdo a los clados formados en la Figura 3.6

Clado 1													
*Contig	Clave	Longitud (nt)	Descripción	Anotación ID	Bit score	Coertura q. (%)	Coertura t. (%)	Identidad (%)	Valor e	SN1 TPM	SN2 TPM	SR1 TPM	SR2 TPM
c9724_6_g1_i1	LHCA 2	1323	Proteína 2 de unión a clorofila del complejo I	cann107838850	566.6	61.3	100	100	6E-159	1929.57	1543.67	898.55	1023.98
c6689_8_g1_i1	KARI	2201	Cetol-ácido reductoiso merasa	cann107856177	1115.5	77.19	95.77	100	0	1440.41	1181.7	1097.56	1052.82
c5178_9_g1_i1	LHCB 3	1217	Proteína 3 de unión a clorofila del complejo II	cann:107878708	539.7	64.67	100	99	7E-151	1186.28	1166.3	454.3	664.09
Clado 2													
*Contig	Clave	Longitud (nt)	Descripción	Anotación ID	Bit score	Coertura q. (%)	Coertura t. (%)	Identidad (%)	Valor e	SN1 TPM	SN2 TPM	SR1 TPM	SR2 TPM
c9779_7_g1_i1	GSA	1887	Glutamato-1-semialdehído 2,1 aminomutasa	cann107863035	940.3	76.05	100	98	3E-271	65.58	49.38	42.98	44.03
c1077_03_g1_i1	OTC	1359	Ornitina transcarbamilasa	cann:107866980	735.3	85.5	97.48	95	1E-209	57.72	45.06	38.81	42.96
c4963_9_g1_i7	PRR3 7	3175	Regulador de pseudoresp	cann107850219	1344.7	73.83	100	97	0	55.58	48.28	46.43	42.94

1			uesta 7										
c8033 4_g1_i 1	PDS	2217	15-cis fitoeno desaturasa	cann107861 625	1164.1	78.8	100	99	0	54.09	45.23	51.56	40.97
c6387 9_g4_i 1	RPN2	3762	Subunidad N2 regulatoria del proteasoma 26S	cann107854 708	1723.4	80.01	100	98	0	69.5	53.94	61.1	57.20
c3746 7_g1_i 1	GOGA T	7328	Glutamato sintasa	cann:107855 963	4224.5	90.45	100	97	0	71.41	48.73	54.09	49.81
c6611 4_g6_i 1	GABA T1	2133	Gamma- aminobutira to transamina sa 1	cann:107862 720	741.1	51.66	71.44	100	3E- 211	63.44	49.39	71.42	61.41
c9878 7_g1_i 1	GluTR	2327	Glutamil- tRNA reductasa	cann107867 325	1007.3	70.69	100	96	2E- 291	136.7 2	80.83	68.25	73.35
c1071 1_g1_i 1	DXS	3107	1-Desoxi- xilulosa-5- fosfato sintasa	cann107852 315	1426.4	69.46	100	99	0	108.1 3	75.5	51.89	52.79
c7319 4_g1_i 1	ARG1	1593	Arginasa 1	cann:107873 005	676	65.22	100	99	8E- 192	88.29	94.66	84.74	75.33
c6206 7_g3_i 1	GS	1506	Glutamina sintetasa	cann107870 519	740	70.98	100	99	4E- 211	126.3 3	118.5 5	108.0 1	107.8 6

ANEXOS

c2410 4_g1_i 1	AK	2452	Aspartato cinasa	cann107838 766	1094	69.05	100	99	0	119.3 3	82.72	100.4 2	94.34
c2734 _g1_i1	6PGD	1965	6- fosfoglucon ato deshidroge nasa	cann:107850 704	980.7	75.62	100	99	2E- 283	165.8 8	132.6 4	124.3	125.6 7

Clado 3

*Conti g	Clave	Longitud (nt)	Descripción	Anotación ID	Bit score	Cober tura q. (%)	Cober tura t. (%)	Identi dad (%)	Valor e	SN1 TPM	SN2 TPM	SR1 TPM	SR2 TPM
c6101 7_g1_i 2	UGTII	1301	UDP- glucosiltran sferasa	cann:107867 844	729.2	81.48	76.91	99	7E- 208	7.60	5.3	5.32	4.85
c6215 4_g1_i 3	PSY2	2040	15-cis fitoeno sintasa 2	cann107859 651	861.3	63.58	100	100	2E- 247	6.61	5.30	4.29	3.87
c6591 3_g4_i 6	MYB	4029	Factor de transcripci on MYB	XM_0167252 42.2	992.8	13.48	54.08	99	0	3.77	2.88	9.79	3.53
c6101 7_g2_i 1	UGTI	1232	UDP- glucosiltran sferasa	cann:107870 274	668.7	81.66	72.51	97	1E- 189	2.27	0.97	2.51	2.08
c5441 6_g1_i 1	LHY	1990	Hipocótilo alargado tardío	cann:107845 265	1068.9	86.28	73.15	97	0	1.37	1.41	1.48	1.61
c6492 3_g1_i 1	AT3D2	1177	AT3-D2	PHU26685.1	703.7	89.29	79.37	98	3E- 199	2.04	0.77	2.40	0.56
c6611 4_g8_i	GABA	1834	Gamma- aminobutira	cann:107841	988	79.39	92.73	99	1E-	12.71	15.34	9.37	18.36

1	T3		to transamina sa 3	786						285			
c7262 1_g1_i 1	DXS2	2585	1-Desoxi- xilulosa-5- fosfato sintasa 2	cann107850 768	1370.9	82.9	100	96	0	42.71	26.58	20.16	27.45

Clado 4

*Conti g	Clave	Longitud (nt)	Descripción	Anotación ID	Bit score	Cober tura q. (%)	Cober tura t. (%)	Identi dad (%)	Valor e	SN1 TPM	SN2 TPM	SR1 TPM	SR2 TPM
c1070 8_g1_i 1	G6PD H	2114	Glucosa-6- fosfato deshidroge nasa	cann:107860 735	1032.7	72.42	100	100	5E- 299	217.2 1	191.6 1	184.5 3	169.7 1
c6720 3_g1_i 1	RPT4	1746	Subunidad T4 regulatoria del proteasoma 26S	cann107840 436	778.9	68.44	100	100	1E- 222	220.2 2	190.1	173.1 1	176.4 7
c6574 4_g2_i 1	KAS	3716	Cetoacil- ACP sintasa	AAC78479.1	1006	39.17	100	100	0.0	175.5 5	148.2	206.5 2	76.67
c8951 3_g1_i 1	FAT	1292	Acil-ACP tioesterasa	AAG35064.1	741.1	85.53	99.19	99	2E- 210	182.4 8	111.6 6	171.2 5	96.13
c9724 2_g1_i 1	KR	1802	Acil-ACP reductasa	PHU15508.1	587.4	53.66	100	96	5E- 164	235.3 9	190.6 6	336.0 9	190.5 2
c6517 2	AT3D1	1712	Capsaicina sintasa,	AAV66308.1	870.5	77.16	100	98	3E- 249	309.7 4	230.8 2	378.7 4	191.7 5

ANEXOS

_g1_i1

PUN1

Clado 5

*Contig	Clave	Longitud (nt)	Descripción	Anotación ID	Bit score	Coertura q. (%)	Coertura t. (%)	Identidad (%)	Valor e	SN1	Clave	Longitud (nt)	Descripción
c6488 1_g10 _i1	pAMT	3340	Aminotransferasa putativa	AAC78480.1	929.9	41.26	100	100	8E-267	516.32	386.1	696.03	337.85

*La identificación de la anotación corresponde a la base de datos de KEGG o del NCBI. La cobertura q. corresponda a la cobertura del *contig* ("query"). La cobertura t. corresponde a la cobertura de la secuencia de referencia de bases de datos "target". La clave de cada gen se usó en el mapa de calor de la Figura 3.6 y en la Tabla 3.3

Tabla Suplementaria 3.3. Contenido de capsaicinoides en frutos de plantas de chile habanero crecidas en suelo negro (SN) o suelo rojo (SR) en distintas etapas de desarrollo postantesis (DPA). Se presentan los resultados como g de capsaicina/g de peso seco de fruto y como unidades de Scoville (SHU)

DPA	25	45	60
SN			
Capsaicina (g/g PS)	0.01072	0.0113	0.00772
SHU*	160800	169500	115800
SR			
Capsaicina (g/g PS)	0.00892	0.01148	0.00987
SHU*	133800	172200	148050

*SHU= (g capsaicina/g PS)(15000000)

Tabla Suplementaria 3.4. Frutos totales de chile habanero de 45 DPA en distintos estados de madurez (completamente maduro, color naranja; en estado de madurez intermedio, en cambio de color; e inmaduros, es decir, de color verde) en suelo negro y suelo rojo

Estado de madurez	Suelo negro (número)	Suelo rojo (número)
Frutos maduros (naranjas)	47 (8.41%)	39 (8.35)
Frutos en estado intermedio	298 (53.31%)	213 (45.61%)
Frutos inmaduro (verdes)	214 (38.28%)	215 (46.04%)
Total	559 (100%)	467 (100%)

Supplementary Table 4.1. AT3 sequences and homologues in the genus *Capsicum* retrieved from NCBI database

Code	Group	Specie	Cultivar	NCBI annotation	ID accession / size (aa)	Exon I (nt), II (nt)	ID nucleotide accession/ size (nt)
can01		<i>C. annuum</i>	Thai Hot	Acyltransferase (Pun1)	AAV66311.1/440	1-738, 1087-1671	AY819029.1 /1671 /1323
can02		<i>C. annuum</i>	Hot 1493	Acyltransferase (Pun1)	AAV66310.1/440	1-738, 1087-1671	AY819028.1 /1671 /1323
can03	Group I (AT3-D1)	<i>C. annuum</i>	Zunla-1	Acylsugar acyltransferase 3-like, AT3, mRNA	NP_001311698.1/440	1-738, 739-1323 1-738, 1087-1671	NM_001324769.1/1323 CM002813.1 /1671
can04		<i>C. annuum</i>	Yujiao5	Acyltransferase 3-like (AT3-D1), mRNA	QCT85208.1/440	1-738, 739-1323	MK138366.1 /1323

ANEXOS

can05	<i>C. annuum</i>	RNaky	Acyltransferase (AT3-1)	ACO48259.1/415	1-738, 1087-1593	FJ755173.1/1593 /1245
can07	<i>C. annuum</i>	Taiwan2	TSA, RNA sequence	Traduced with ORFFinder/40	1-738, 739-1323	GAXU01020 367.1/1323
can08	<i>C. annuum</i>	Pusa Jwala	TSA, RNA sequence	Traduced with ORFFinder/40	1-738, 739-1323	GAXV01019 436.1/1323
can09	<i>C. annuum</i>	-----	TSA, RNA sequence	Traduced with ORFFinder/40	1-738, 739-1323	JW074441.1 /1323
can10	<i>C. annuum</i>	277long	Acyltransferase, caft1 gene	BAD99417.1/405	1-634, 982-1566	AB206919.1 /1566 /1219
can11	<i>C. annuum</i>	754	Acyltransferase 3-like (AT3-D1) mRNA	QCT85209.1/440	1-738, 739-1323	MK138367.1 /1323
can13	<i>C. annuum</i>	Takanotsu me	Acyltransferase Pun1 gene	BBG28791.1/440	1-738, 1087-1671	LC423556.1/1671 /1323
can15	<i>C. annuum</i>	SF	Sequence from genome search	Traduced with ORFFinder/40	1-738, 1087-1671	VYZZ01006 408.1/1671 /1323
can19	<i>C. annuum</i>	CM334	Sequence from genome search	Traduced with ORFFinder/40	1-738, 1087-1671	CM008456.1 /1671 /1323
can22	<i>C. annuum</i>	Glabriusculum	Sequence from genome search	Traduced with ORFFinder/40	1-738, 1087-1671	CM003180.1 /1671 /1323
*can25	<i>C. annuum</i>	Jupiter	pun1 allele	-----	1-49, 398-982	AY819030.1 /982
*can26	<i>C. annuum</i>	Maor	pun1 allele	-----	1-49, 398-982	AY819031.1 /982
*can27	<i>C. annuum</i>	Sweet3575	pun1 allele	-----	1-49, 398-982	AY819032.1 /982

*can3 3	<i>C. annuum</i>	ECW	Sequence from genome search	-----	1-49, 398-982	VYZY01009 709.1/982
*can3 7	<i>C. annuum</i>	UCD-10X-F1	Sequence from genome search	-----	1-49, 398-982	CM009387.1 /982
cch01	<i>C. chinense</i>	Habanero	Acyltransferase (Pun1) gene	AAV66309.1/ 440	1-738, 1087- 1671	AY819027.1 /1671 /1323
cch02	<i>C. chinense</i>	PI159236	Hypothetical protein/transferase	PHU26684.1/ 440	1-738, 1087- 1671	CM008432.1 /1671 /1323
*cch0 5	<i>C. chinense</i>	NMCA300 36	Acyltransferase pun1 ² allele	-----	1-734, 1083- 1667	EF104910.1/ 1667
cfr01	<i>C. frutescens</i>	BG2814-6	Acyltransferase (Pun1) gene	AAV66308.1/ 440	1-738, 1087- 1671	AY819026.1 /1671 /1323
cfr02	<i>C. frutescens</i>	Tabasco	Acyltransferase (Pun1) gene	ACO48260.1/ 415	1-738, 1087- 1593	FJ755174.1/ 1593 /1245
cfr03	<i>C. frutescens</i>	PI594141- P	Acyltransferase (Pun1) gene	ACO48261.1/ 415	1-738, 1087- 1593	FJ755175.1/ 1593 /1245
cfr04	<i>C. frutescens</i>	Shuanla	Acyltransferase (Pun1) gene	ADN97116.1/ 440	1-738, 1087- 1671	HM854860.1 /1671 /1323
*cfr05	<i>C. frutescens</i>	PI594141- np	Acyltransferase pun1 ³ allele	-----	1-738, 1088- 1480	FJ871985.1/ 1480
cba01	<i>C. baccatum</i>	PBC81	Hypothetical protein/transferase	PHT56199.1/ 440	1-738, 1090- 1674	MLFT02000 002.1/1674 /1323
cha01	<i>C. chacoense</i>	-----	Acyltransferase (AT3-1) gene	ACO48262.1/ 415	1-738, 1123- 1629	FJ755176.1/ 1629 /1245
cha02	<i>C. chacoense</i>	PI260433	Acyltransferase (AT3) gene	ACZ28491.1/ 415	1-738, 1122- 1628	FJ871984.1/ 1628 /1245
can06	<u>Group II (AT3-D2)</u> <i>C. annuum</i>	Zunla-1	Acylsugar acyltransferase 3-like, mRNA	XP_0165602 62.1/441	1-738, 739-1326, 1-738,	XM_016704 776.1/1326, CM002813.1

ANEXOS

						1409-1996	/1996
can12	<i>C. annuum</i>	754	Acyltransferase 3-like (AT3-D2) mRNA	QCT85207.1/441		1-738, 739-1326	MK138365.1/1326
can14	<i>C. annuum</i>	ECW	Sequence from genome search/deacetyl vindoline-O-acetyltransferase-like	KAF3623437.1/441		1-738, 1394-1981	VYZY01009709.1/1981/1326
can16	<i>C. annuum</i>	SF	Sequence from genome search/deacetyl vindoline-O-acetyltransferase	KAF3637602.1/441		1-738, 1409-1996	VYZZ01006408.1/1996/1326
can18	<i>C. annuum</i>	UCD-10X-F1	Sequence from genome search	Traduced with ORFFinder/441		1-738, 1394-1981	CM009387.1/1981/1326
can20	<i>C. annuum</i>	CM334	Sequence from genome search	Traduced with ORFFinder/441		1-738, 1409-1996	CM008456.1/1996/1326
can23	<i>C. annuum</i>	Glabriusculum	Sequence from genome search	Traduced with ORFFinder/441		1-738, 1409-1996	CM003180.1/1996/1326
cch03	<i>C. chinense</i>	PI159236	Hypothetical protein/transferase	PHU26685.1/441		1-738, 1409-1996	MCIT02000002.1/1996/1326
cba02	<i>C. baccatum</i>	PBC81	Hypothetical protein/transferase	PHT26523.1/442		1-741, 1411-1998	MLFT02000752.1/1998/1329
*can17	<i>C. annuum</i>	SF	Sequence from genome search/3beta-hydroxysteroid-dehydrogenase/descarboxylase isoform 2-like	KAF3637600.1	Group III	1-750, 961-1509	VYZZ01006408.1/1509/1299
*can21	<i>C. annuum</i>	CM334	Sequence from genome search	Traduced with ORF		1-750, 961-1509	CM008456.1/1509/1299

				Finder/432		
*can2 4	<i>C. annuum</i>	Glabriusculum	Sequence from genome search	Traduced with ORF Finder/432	1-750, 961-1509	CM003180.1 /1509 /1299
*can3 2	<i>C. annuum</i>	Zunla-1	Sequence from genome search	Traduced with ORF Finder/385	1-750, 961-1368	CM002813.1 /1368 /1158
*can3 4	<i>C. annuum</i>	ECW	Sequence from genome search	-----	1-748, 959-1477	VYZY01009 709.1/1477
*can3 8	<i>C. annuum</i>	UCD-10X-F1	Sequence from genome search	-----	1-748, 959-1555	CM009387.1 /1555
*cch0 4	<i>C. chinense</i>	PI159236	Sequence from genome search	-----	1-751, 962-1558	CM008432.1 /1558
*crh01	<i>C. rhomboideum</i>	-----	Acyltransferase pseudogene	-----	1-738, 1091-1618	MF142764.1 /1618
*crh02	<i>C. rhomboideum</i>	-----	Acyltransferase pseudogene (AT3-1/AT3-2 fusion)	-----	1-738, 1091-1617	FJ755165.1/1617
*crh03	<i>C. rhomboideum</i>	-----	Pseudogene AT3-2	-----	1-752, 1153-1611	FJ687529.1/1611
*can2 8	<i>C. annuum</i>	Thai Hot	Pseudogene AT3-2	-----	1-732-1133-1591	FJ687530.1/1591
*can2 9	Group IV, pseudo genes <i>C. annuum</i>	RNaky	Pseudogene AT3-2	-----	1-733, 1134-1591	FJ687524.1/1591
*can3 0	<i>C. annuum</i>	Maor	Pseudogene AT3-2	-----	1-751, 1152-1611	FJ755161.1/1611
*can3 1	<i>C. annuum</i>	Zunla-1	Sequence from genome search	-----	1-733, 1134-1713	CM002813.1 /1713
*can3 5	<i>C. annuum</i>	ECW	Sequence from genome search	-----	1-752, 1153-1746	VYZY01009 709.1/1746
*can3 6	<i>C. annuum</i>	SF	Sequence from genome search	-----	1-733, 1134-	VYZZ01006 408.1/1713

ANEXOS

					1713	
*can39	<i>C. annuum</i>	UCD-10-F1	Sequence from genome search	-----	1-752, 1153-1746	CM009387.1 /1746
*can40	<i>C. annuum</i>	CM334	Sequence from genome search	-----	1-733, 1134-1713	CM008456.1 /1713
*can41	<i>C. annuum</i>	Glabriusculum	Sequence from genome search	-----	1-756, 1156-1740	CM003180.1 /1740
*cch06	<i>C. chinense</i>	Habanero	Pseudogene AT3-2	-----	1-750, 1177-1638	FJ755163.1 /1638
*cch07	<i>C. chinense</i>	PI159236	Sequence from genome search	-----	1-751, 1178-1774	CM008432.1 /1774
*cfr06	<i>C. frutescens</i>	BG2814-6	Pseudogene AT3-2	-----	1-754, 1155-1748	FJ755162.1 /1748
*cfr07	<i>C. frutescens</i>	BG2814-6	Pseudogene AT3-2	-----	1-501, 902-1360	FJ687526.1 /1360
*cfr08	<i>C. frutescens</i>	PI594141	Pseudogene AT3-2	-----	1-722, 1123-1581	FJ687527.1 /1581
*cfr09	<i>C. frutescens</i>	Tabasco	Pseudogene AT3-2	-----	1-747, 1147-1605	FJ755160.1 /1605
*cba03	<i>C. baccatum</i>	PBC81	Sequence from genome search	-----	1-750, 1153-1743	CM008444.1 /1743
*cha03	<i>C. chacoense</i>	-----	Pseudogene AT3-2	-----	1-751, 1152-1611	FJ755164.1 /1611

*Sequences discarded to haplotype analysis

Supplementary Table 4.2. Details of the regulatory regions at 2000 pb upstream of the star of *AT3-D1* and *AT3-D2* on *C. annuum* Zunla-1 (CM002813.1) and *C. chinense* PI159236 (CM008432.1)

Factor	Possible function	Reference plant	<i>AT3-D1</i> (cch02)	<i>AT3-D2</i> (cch03)	<i>AT3-D1</i> (can03)	<i>AT3-D2</i> (can06)	Sequence
AAGAA motif		<i>Avena sativa</i>	-1589(-) -698(-)	-1510(-)	-1590(-)	-1510(-)	gGTAAAGA AA GAAAGAA
ABRE	Cis-acting element involved in the abscisic acid responsiveness	<i>Arabidopsis thaliana</i>	-365(-) -364(+)	-1500(-) -540(-) -941(-) -539(+)	-365(-) -364(+)	-539(-) -1500(-) -944(-)	CACGTG ACGTG
ABRE3a	Cis-acting element involved in the abscisic acid responsiveness	<i>Zea mays</i>		-941(-)		-944(-)	TACGTG
ABRE4	Cis-acting element involved in the abscisic acid responsiveness	<i>Zea mays</i>		-941(+)		-944(+)	CACGTA
ACE	Cis-acting element involved in light responsiveness	<i>Petroselinum crispum</i>		-1301(-)		-1304(-)	CTAACGTA TT
AE-Box	Part of a module for light response	<i>Arabidopsis thaliana</i>		-837(+)		-837(+)	AGAAACTT
ARE	Cis-acting element essential for the anaerobic induction	<i>Zea mays</i>	-1119(+) -1108(+)		-1120(+) -1109(+)		AAACCA

ANEXOS

AT-rich element	Binding site of AT-rich DNA binding protein (ATBP-1)	<i>Glycine max</i>	-1273(-)	-1449(-)	-1274[(-)	-1451(-)	ATAGAAAT CAA
ATCT-motif	Part of a conserved DNA module involved in light responsiveness	<i>Pisum sativum</i>				-920(-)	AATCTAATC C
AT-TATA-box	Core promotor element	<i>Arabidopsis thaliana</i>	-186(-)	-1703(+)	-578(-)	-1703(+)	TATATA
			-72(-)	-73(-)	-186(-)	-1701(+)	TATATAAA
				-1701(+)	-576(-)		
					-72(-)		
AuxRE	Part of an auxin-response element	<i>Glycine max</i>				-1672(-)	TGTCTCAAT AAG
Box 4	Part of a conserved DNA module involved in light responsiveness	<i>Petroselinum crispum</i>	-1337(+)	-1972(+)	-1338(+)	-1972(+)	ATTAAT
			-717(-)	-7(-)	-406(-)	-7(-)	
			-406(-)	-1122(+)	-717(-)	-1125(+)	
			-7(-)	-1783(+)	-7(-)	-581(-)	
				-582(-)		-1783(+)	

CAAT- Box	Common cis-acting element in promoter and enhancer regions	<i>Nicotiana glutinosa</i>	-1972(+)	-1990(-)	-1973(+)	-1990(-)	CAAT
			-1880(+)	-1877(+)	-1881(+)	-1804(+)	
			-1682(+)	-1839(-)	-1683(+)	-1877(+)	CAAAT
		<i>Pisum sativum</i>	-1498(+)	-1804(+)	-1569(+)	-1686(+)	
			-1425(+)	-1736(+)	-1499(+)	-1672(+)	CCAAT
		<i>Arabidopsis thaliana</i>	-1399(+)	-1723(-)	-1426(+)	-1736(+)	
			-1383(+)	-1686(+)	-1400(+)	-1839(-)	
			-1327(+)	-1629(+)	-1384(+)	-1629(+)	
			-1309(+)	-1588(-)	-1328(+)	-1723(-)	
			-1274(-)	-1536(-)	-1310(+)	-1588(-)	
			-1266(+)	-1287(-)	-1275(-)	-1546(-)	
			-1062(-)	-1205(+)	-1267(+)	-1363(+)	
			-903(-)	-1196(+)	-1245(-)	-1208(+)	
			-690(+)	-1194(-)	-1063(-)	-1199(+)	
			-616(-)	-1043(+)	-903(-)	-1197(-)	
			-554(+)	-1011(-)	-690(+)	-1046(+)	
			-524(-)	-971(+)	-616(-)	-1014(-)	
			-380(+)	-564(-)	-554(+)	-974(+)	
			-308(+)	-534(+)	-524(-)	-563(-)	
			-298(-)	-479(-)	-380(+)	-533(+)	
				-468(+)	-308(+)	-478(-)	
			-1550(+)	-218(+)	-298(-)	-467(+)	
			-1491(+)	-25(-)		-217(+)	
			-1464(+)		-1551(+)	-25(-)	
			-1187(+)	-1745(+)	-1492(+)		
			-1096(+)	-1678(-)	-1465(+)	-1742(-)	
			-830(-)	-1609(-)	-1188(+)	-1627(-)	

ANEXOS

			-726(-)	-1467(+)	-1097(+)	-1047(+)	
			-720(+)	-1386(-)	-830(-)	-218(+)	
			-484(-)	-1361(+)	-726(-)		
			-426(+)	-1185(-)	-720(+)	-1678(-)	
			-421(-)	-1038(-)	-433(-)	-1745(+)	
			-16(-)	-930(-)	-426(+)	-1909(-)	
			-868(-)	-763(-)	-421(-)	-1609(-)	
				-653(-)	-16(-)	-1469(+)	
			-859(-)	-608(-)		-1388(-)	
			-805(-)	-602(+)	-868(-)	-1188(-)	
			-691(+)	-597(-)	-859(-)	-933(-)	
				-567(+)	-805(-)	-763(-)	
				-16(-)	-691(+)	-653(-)	
					-483(-)	-608(-)	
					-1742(-)	-601(+)	
					-1627(-)	-596(-)	
					-1044(+)	-566(+)	
					-219(+)	-16(-)	
CGTCA- motif	Cis-acting regulatory element involved in the MeJa- responsive ness	<i>Hordeum vulgare</i>	-1883(+)		-1884(+)		CGTCA
CAT- Box	Cis acting regulatory element related to meristem expression	<i>Arabidopsis thaliana</i>		-1820(+)		-1820(+)	GCCACT
ERE	Ethylene- response	<i>Nicotiana glutinosa</i>	-1291(+)		-1292(+)		ATTCATA

element			-1249(+)				
F-Box	F-Box element. Involved in many processes, like the regulation of the floral development	<i>Solanum lycopersicum</i>	-672(+)		-672(+)		CTATTCTCA TT
G-Box	Cis-acting regulatory element involved in light responsive ness	<i>Pisum sativum</i>	-365(-)	-540(-)	-365(-)	-539(-)	CACGTG
				-1812(-)		-1812(-)	
		<i>Arabidopsis thaliana</i>		-941(-)		-1500(+)	CACGAC
				-1500(+)		-944(-)	
		<i>Zea mays</i>					TACGTG
GA-motif	Part of a light responsive element	<i>Arabidopsis thaliana</i>		-1594(-)		-1594(-)	ATAGATAA
GARE-motif	Gibberellin-responsive element	<i>Brassica oleracea</i>	-1512(-)		1513(-)		TCTGTTG
GATA-motif	Part of a light responsive element	<i>Arabidopsis thaliana</i>	-1983(-)	-1416(-)	1984(-)	-1418(-)	AAGATAAG ATT
						-340(+)	
							GATAGGA
							GATAGGG
GT1-motif	Light responsive element	<i>Arabidopsis thaliana</i>	-247(-)	-1570(+)	-247(-)	-1570(+)	GGTTAA
				-437(-)		-436(-)	
Gap-Box	Part of a light responsive element	<i>Arabidopsis thaliana</i>	-1491(+)		-1492(+)		CAAATGAA(A/G)A

ANEXOS

MBS	MYB binding site involved in drought-inducibility	<i>Arabidopsis thaliana</i>	-1655(+)	-1145(-)	-1656(+)		CAACTG
MRE	MYB binding site involved in light responsiveness	<i>Petroselinum crispum</i>		-1064(+)		-1067(+)	AACCTAA
MBSI	MYB binding site involved in flavonoid biosynthetic genes regulation	<i>Petunia hybrida</i>	-1049(+)				aaaAaaC(G/C)GTTA
MYB	Binding site of MYB transcription factors, that can be involved in responsiveness to abiotic stress	<i>Arabidopsis thaliana</i>	-1931(+)	-1658(-)	-1513(+)	-1658(-)	CAACCA
			-1512(+)	-402(+)	-1656(+)	-401(+)	
		<i>Nicotiana tabacum</i>	-1655(+)	-1145(-)		-1078(-)	CAACTG
						CAACAG	
						TAACCA	
MYC	Can be involved in the response to abiotic stress, like drought	<i>Arabidopsis thaliana</i>	-1491(-)	-1629(+)	-1492(-)	-1629(+)	CATTTG
			-320(-)	-1196(+)	-320(-)	-1199(+)	CATGTG
			-1464(-)		-1465(-)		CAATTG
P-Box	Gibberellin-responsive element	<i>Oryza sativa</i>	-81(+)	-82(+)	-81(+)	-82(+)	CCTTTTG
O2-site	Cis-acting regulatory element involved in zein metabolism regulation	<i>Zea mays</i>				-1578(+)	GATGATGTGG

STRE	Stress response element	<i>Arabidopsis thaliana</i>	-980(-)	-1487(-)	980(-)	-1488(-)	AGGGG
				-118(+)		-178(+)	
				-195(+)		-373(+)	
				-374(+)		-118(+)	
				-178(+)			
Pc-CMACc	Part of a light responsive element	<i>Pisum sativum</i>			-788(-)		GCCCACGC A
TATA-Box	Core promoter element	<i>Arabidopsis thaliana</i>	-875(-)	-1978(-)	-875(-)	-1978(-)	TATTTAAA
			-72(-)	-1977(+)	-839(-)	-1915(-)	
		<i>Brassica napus</i>	-186(-)	-1915(-)	-656(-)	-1437(-)	TATATA
			-56(-)	-1914(+)	-651(-)	-1213(-)	
			-187(-)	-1801(+)	-644(-)	-1101(-)	TATA
		<i>Brassica oleracea</i>	-70(-)	-1703(+)	-609(-)	-1977(+)	
			-184(-)	-1701(+)	-838(-)	-1946(+)	TATATAA
		<i>Pisum sativum</i>	-23(-)	-1699(+)	-655(-)	-1914(+)	
			-136(-)	-1604(+)	-650(-)	-1801(+)	TATAA
			-839(-)	-1438(-)	-643(-)	-1699(+)	
		<i>Helianthus annuus</i>	-838(-)	-1435(-)	-638(-)	-1604(+)	taTATAAAtc
			-656(-)	-1434(+)	-608(-)	-1436(+)	
		<i>Avena sativa</i>	-655(-)	-1380(-)	-574(-)	-1344(+)	TAAAGATT
			-651(-)	-1341(+)	-377(-)	-1320(+)	
			-650(-)	-1317(+)	-281(-)	-1212(+)	ccTATAAAaa
			-644(-)	-1210(-)	-200(-)	-1191(+)	
			-643(-)	-1209(+)	-184(-)	-1100(+)	ATTATA
			-638(-)	-1188(+)	-136(-)	-804(-)	
			-609(-)	-1101(-)	-70(-)	-496(-)	ATATAT
			-608(-)	-1098(-)	-56(-)	-464(-)	

-577(-)	-1097(+)	-23(-)	-389(-)	ATATAA
-576(-)	-804(-)	-578(-)	-137(-)	
-377(-)	-397(-)	-577(-)	-73(-)	TATAAAA
-281(-)	-465(-)	-187(-)	-57(-)	
-200(-)	-390(-)	-576(-)	-28(-)	TATAAA
-190(-)	-137(-)	-186(-)	-19(-)	
	-73(-)	-72(-)	-1703(+)	TATACA
-840(+)	-71(-)	-190(-)	-1701(+)	
-24(-)	-57(-)		-1440(-)	TATATTTAT ATTT
-188(+)	-28(-)	-840(+)	-1104(-)	
-57(-)	-19(-)	-652(+)	-1382(-)	
-652(+)		-645(+)		
-645(+)	-1916(+)	-188(+)	-1947(+)	
-282(-)	-1702(+)	-282(-)	-1916(+)	
	-1700(+)	-57(-)	-1102(+)	
-71(+)	-1318(+)	-24(-)	-1702(+)	
-185(+)	-1189(+)		-1700(+)	
-378(+)	-1099(+)	-611(-)	-1321(+)	
	-805(-)		-1192(+)	
-611(-)	-466(-)	-610(-)	-805(-)	
	-58(-)		-465(-)	
	-29(-)	-378(+)	-58(-)	
-610(-)	-20(-)	-185(+)	-29(-)	
-578(-)		-71(+)	-20(-)	
	-1802(+)			
	-498(+)		-1802(+)	
	-72(+)		-497(+)	

			-1437(-)		-1606(-)	
			-1212(-)		-1438(-)	
					-1214(-)	
			-1606(-)			
			-1436(-)		-1439(-)	
			-1211(-)		-1215(-)	
			-1395(+)		-1397(+)	
TATC-Box	Cis-acting element involved in gibberellin-responsiveness	<i>Oryza sativa</i>	-1150(+)		-1151(+)	TATCCCA
TCA	Related to the response to salicylic acid	<i>Pisum sativum</i>	-952(-)	-641(+)	-641(+)	TCATCTTCA T
		<i>Nicotiana tabacum</i>		-1996(-)	-1996(-)	CCATCTTTT T
TCT-motif	Part of a light responsive element	<i>Arabidopsis thaliana</i>	-295(-)	-476(-)	-1582(-)	TCTTAC
					-475(-)	
TGA-element	auxin-responsive element	<i>Brassica oleracea</i>	-1196(-)			AACGAC
TGACG-motif	Cis-acting regulatory element involved in MeJa-responsiveness	<i>Hordeum vulgare</i>	-1883(-)		-1884(-)	TGACG
Unnamed 1		<i>Zea mays</i>	-1853(-)		-1854(-)	CGTGG

ANEXOS

			-786(+)		-786(+)		
			-1239(-)		-1240(-)		
			-1855(+)	-1490(+)	-1856(+)	-1490(+)	CTCC
			-934(-)	-274(-)	-913(-)	-273(-)	
			-1545(+)	-1089(-)	-1546(+)	-428(+)	
			-708(-)	-197(-)	-239(+)	-196(-)	
Unnamed 4		<i>Petroselinum hortense</i>	-1764(+)	-1370(+)	-1765(+)	-1174(-)	
			-895(-)	-221(+)	-856(-)	-220(+)	
			-967(-)	-429(+)	-967(-)	-289(-)	
			-239(+)	-1424(+)	-895(-)	-1426(+)	
			-1805(+)	-1171(-)	-1806(+)	-1092(-)	
			-913(-)		-708(-)		
			-1720(+)		-1721(+)		
			-856(-)				
Unnamed 6		<i>Zea mays</i>		-1801(+)			taTAAATATct
W-Box	Binding site of transcription factors WRKY	<i>Arabidopsis thaliana</i>	-1964(-)	-1739(-)	-1965(-)	-1739(-)	TTGACC
			-796(+)		-1244(+)	-322(+)	
WRE3	Response element to mechanical wound	<i>Pisum sativum</i>		-1094(-)		-1097(-)	CCACCT
as-1	Element of responsiveness to salicylic acid and systemic acquired resistance (SAR) and genes related to pathogenes	<i>Arabidopsis thaliana</i>	-1883(-)		1884(-)		TGACG

	is sensitive to salicylic acid				
Circadian	Cis-acting regulatory element involved in circadian control	<i>Solanum lycopersicum</i>	-1911(-)	-1912(-)	CAAAGATA TC

Supplementary Table 4.3. Top 10 structural analogs by I-TASSER with the closest structural similarity to AT3-D1 and AT3-D2

	PBD Hit	Specie	Protein	TM-score	RMSD(Å) ^a	Identity ^b	Coverage ^c
	2bghA	<i>Rauvolfia serpentina</i>	Vinorine synthase	0.907	0.66	0.291	0.914
	4ke4A	<i>Sorghum bicolor</i>	Hydrocinnamoyltransferase	0.811	3.60	0.189	0.927
	4g0bA	<i>Coffea canephora</i>	Hydrocinnamoyltransferase	0.795	3.37	0.207	0.895
	3b2sA	<i>Fusarium graminearum</i>	TRI101	0.786	3.36	0.138	0.886
	2xr7A	<i>Nicotiana tabacum</i>	Malonyltransferase	0.780	3.55	0.172	0.902
AT3-D1	6dd2A	<i>Selaginella moellendorffii</i>	Hydrocinnamoyltransferase	0.778	3.17	0.196	0.871
	2e1tA	<i>Chrysanthemum morifolium</i>	Anthocyanin malonyltransferase	0.773	3.70	0.197	0.904
	6mk2A	<i>Plectranthus scutellarioides</i>	Rosmarinic acid synthase	0.751	3.58	0.221	0.857
	2zbaD	<i>Fusarium sporotrichoides</i>	TRI101	0.704	2.97	0.146	0.779
	3fp0A	<i>Fusarium sporotrichoides</i>	TRI3	0.698	4.43	0.092	0.866
AT3-D2	4ke4A	<i>Sorghum bicolor</i>	Hydrocinnamoyltransferase	0.921	1.65	0.193	0.952
	4g0bA	<i>Coffea</i>	Hydrocinnamoyltransferase	0.920	1.25	0.208	0.939

ANEXOS

	<i>canephora</i>	yltransferase				
6dd2A	<i>Selaginella moellendorffii</i>	Hydrocinnamoyltransferase	0.886	1.66	0.198	0.916
6mk2A	<i>Plectranthus scutellarioides</i>	Rosmarinic acid synthase	0.863	1.77	0.201	0.902
3b2sA	<i>Fusarium graminearum</i>	TRI101	0.815	2.98	0.129	0.912
2xr7A	<i>Nicotiana tabacum</i>	Malonyltransferase	0.798	3.25	0.187	0.900
2bghA	<i>Rauvolfia serpentina</i>	Vinorine synthase	0.785	3.24	0.268	0.878
2e1tA	<i>Chrysanthemum x morifolium</i>	Anthocyanin malonyltransferase	0.784	3.35	0.187	0.887
3fp0A	<i>Fusarium sporotrichoides</i>	TRI3	0.735	4.01	0.084	0.891
2zbaD	<i>Fusarium sporotrichoides</i>	TRI101	0.733	2.71	0.132	0.807

a RMSD between residues that are structurally aligned by TM-align

b Sequence identity in the structurally aligned region

c Coverage of the alignment by TM-align. It is equal to the number of structurally aligned residues divided by the length of the query protein

Supplementary Table 4.4. Pocket properties of the solvent channel of the AT3 proteins

Pocket size	AT3-D1	AT3-D2
Volume (Å ³)	2303.74	1936.78
Surface (Å ²)	2755.27	2390.02
Depth (Å)	36.79	39.08
Aminoacidic composition		
Apolar aminoacid ratio	0.48	0.52
Polar aminoacid ratio	0.33	0.28
Negative aminoacid ratio	0.11	0.07
Positive aminoacid ratio	0.08	0.12

FIGURAS SUPLEMENTARIAS

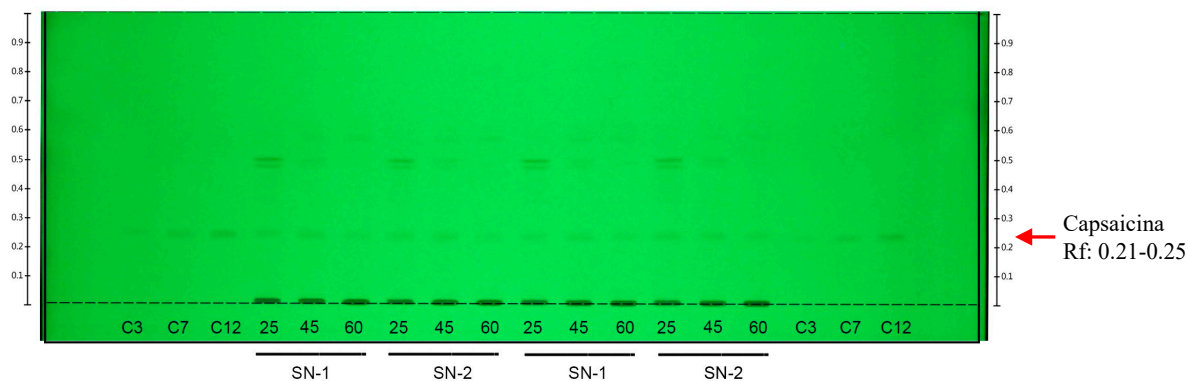


Figura Suplementaria 3.1. Corrida de muestras de capsaicina en frutos de chile habanero por HPTLC. Se realizó la corrida en una placa de sílica gel 60F 254 con una fase móvil ciclohexano:clorofomo:ácido acético (70:20:10) en una cámara CAMAG cerrada sin saturación. Se corrieron muestras de las dos réplicas de frutos de suelo negro (SN-1 y SN-2) junto a estándares de capsaicina de 0.75 μg (C3), 1.75 μg (C7) y 3 μg (C12)

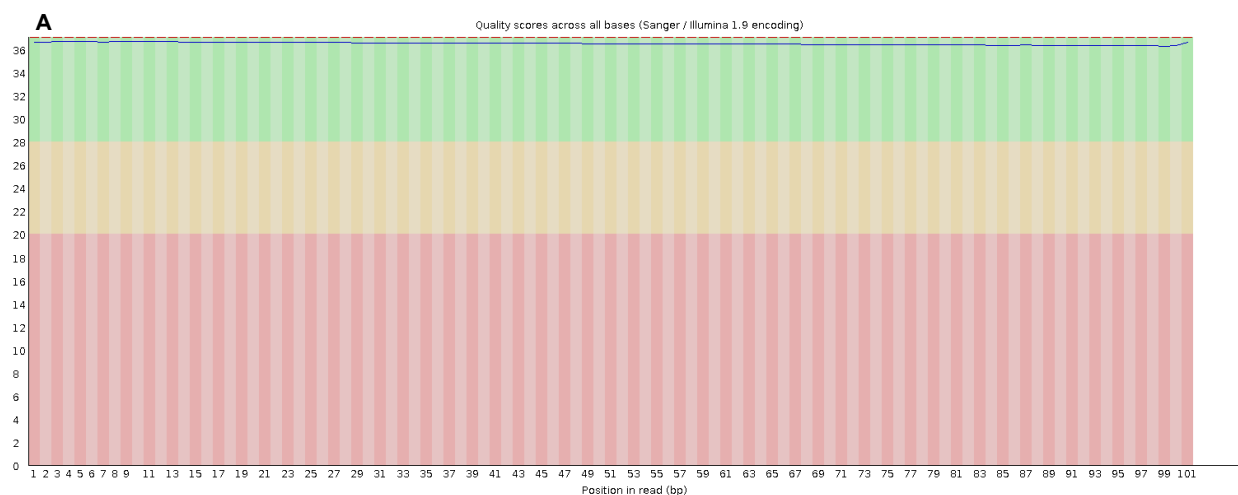
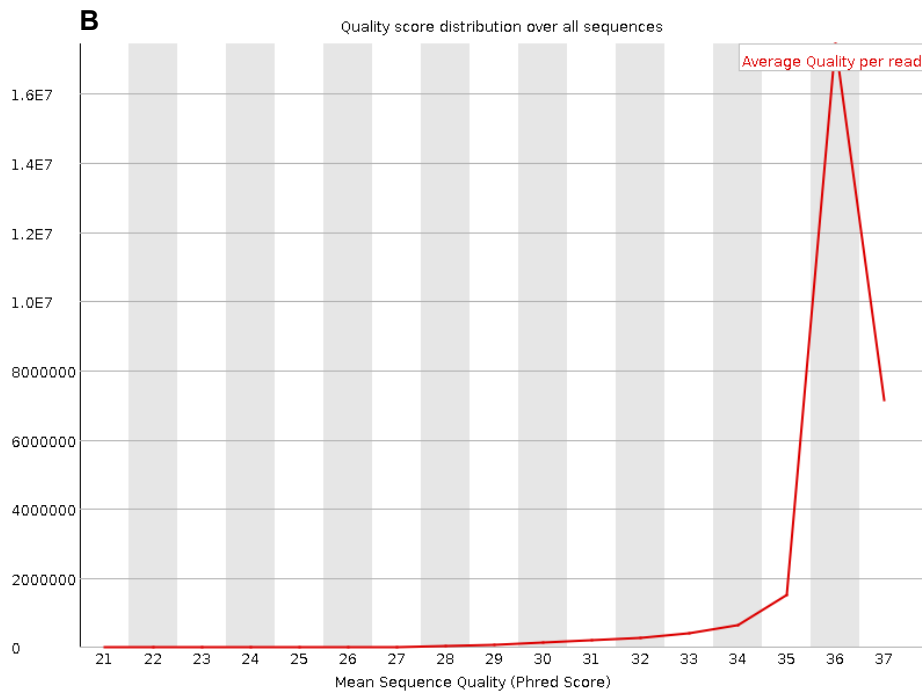
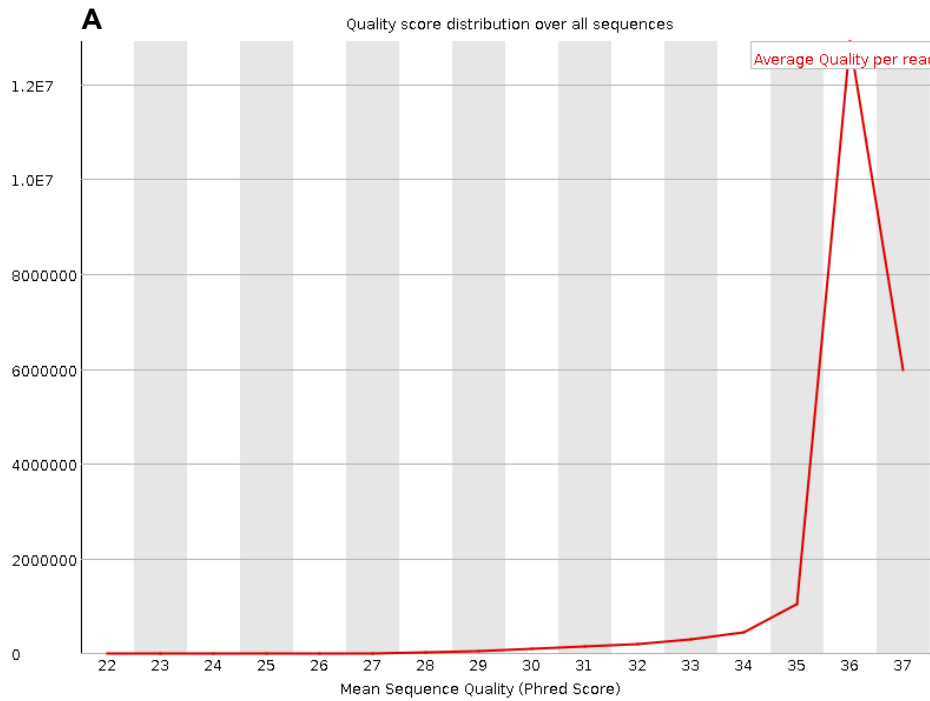




Figura Suplementaria 3.2. Calidad por base en las secuencias (*reads*) en el transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro (SN) y suelo rojo (SR) tras el *trimming* y filtrado de adaptadores, *reads* y bases de baja calidad. Cada transcriptoma tiene dos

réplicas. A) SN1, B) SN2, C) SR1, D) SR2. El eje y muestra los valores de calidad (phred score). El color en la gráfica divide la calidad en tres rangos o zonas: muy buena calidad (verde), calidad razonable (naranja) y baja calidad (rojo). La línea azul representa el promedio de la calidad de todos los *reads*. Las Figuras son representativas de las secuencias forward de las bibliotecas



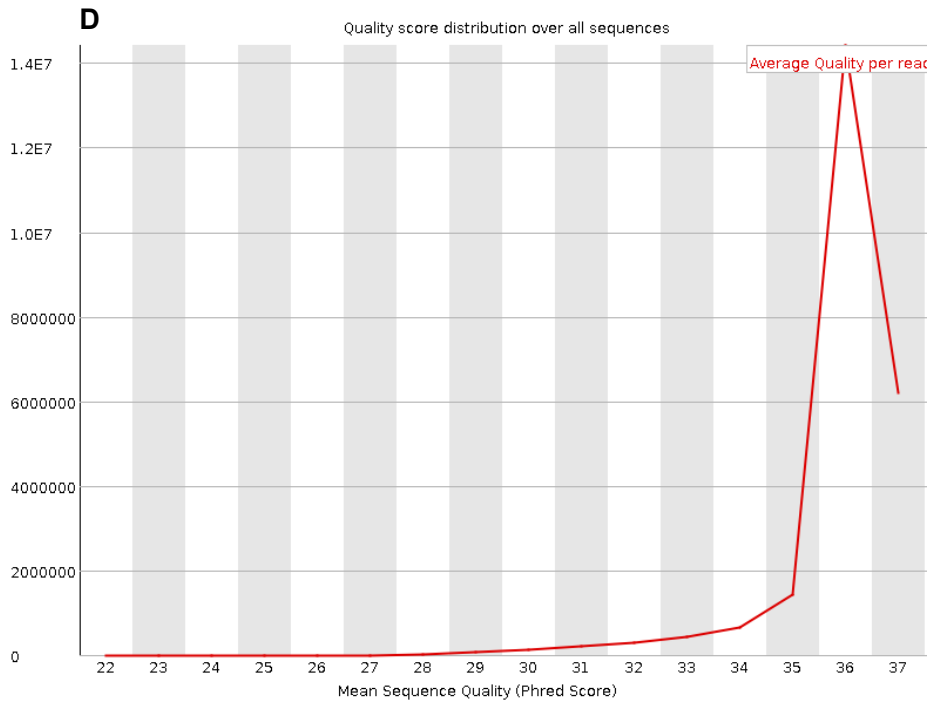
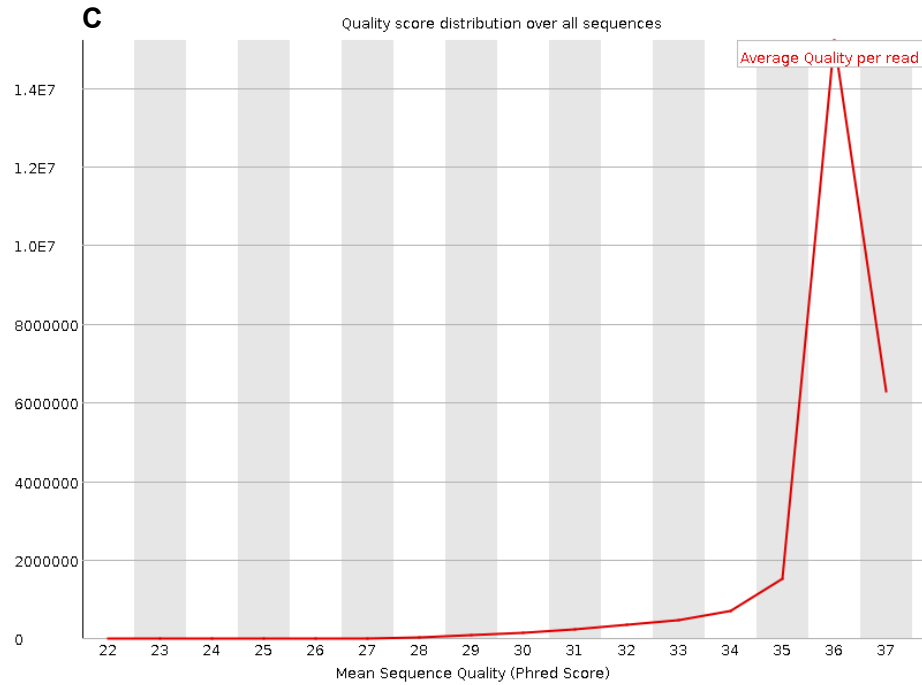
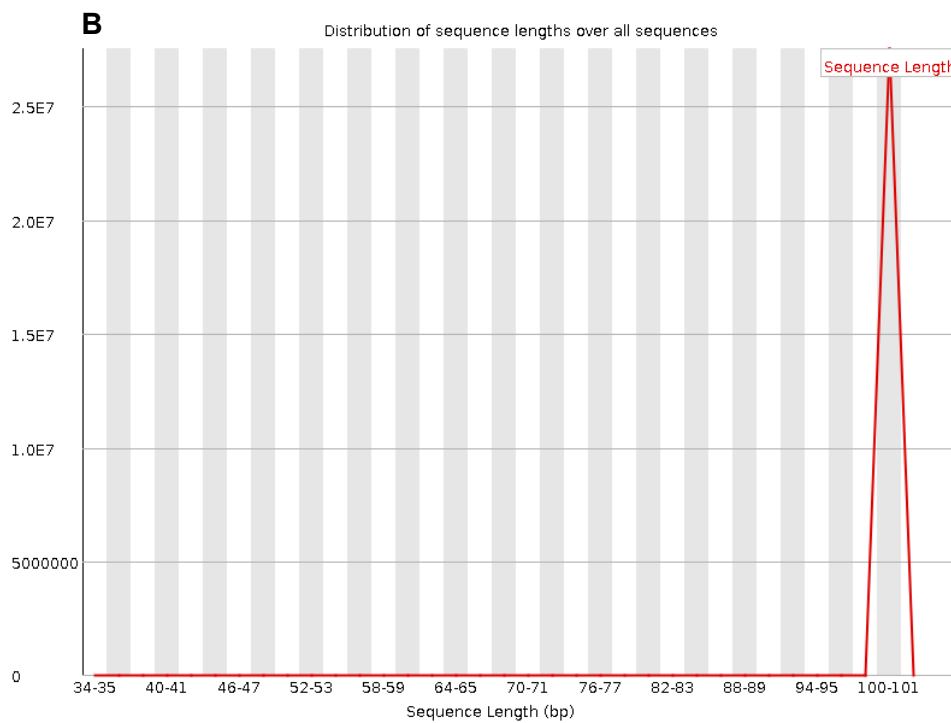
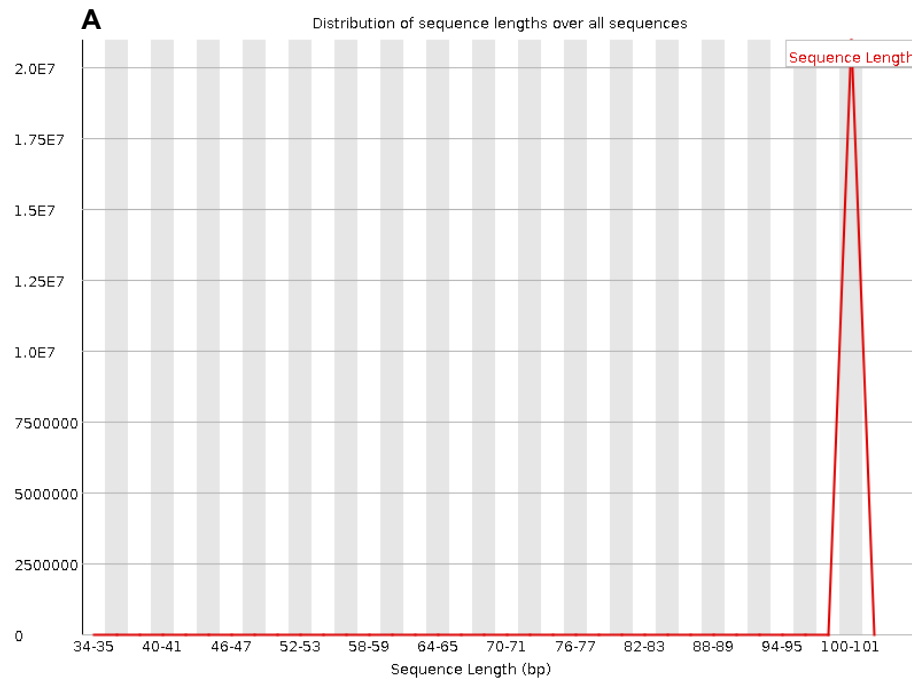


Figura Suplementaria 3.3. Distribución de la calidad en las secuencias (*reads*) del transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro (SN) y suelo rojo (SR) tras el *trimming* y filtrado de adaptadores, *reads* y bases de baja calidad. Cada transcriptoma tiene dos réplicas. A) SN1, B) SN2, C) SR1, D) SR2. El eje x representa la calidad (valores phred) y el eje

y el número de secuencias. El pico en la gráfica señala la calidad promedio máxima de la biblioteca. Las Figuras son representativas de las secuencias forward de las bibliotecas



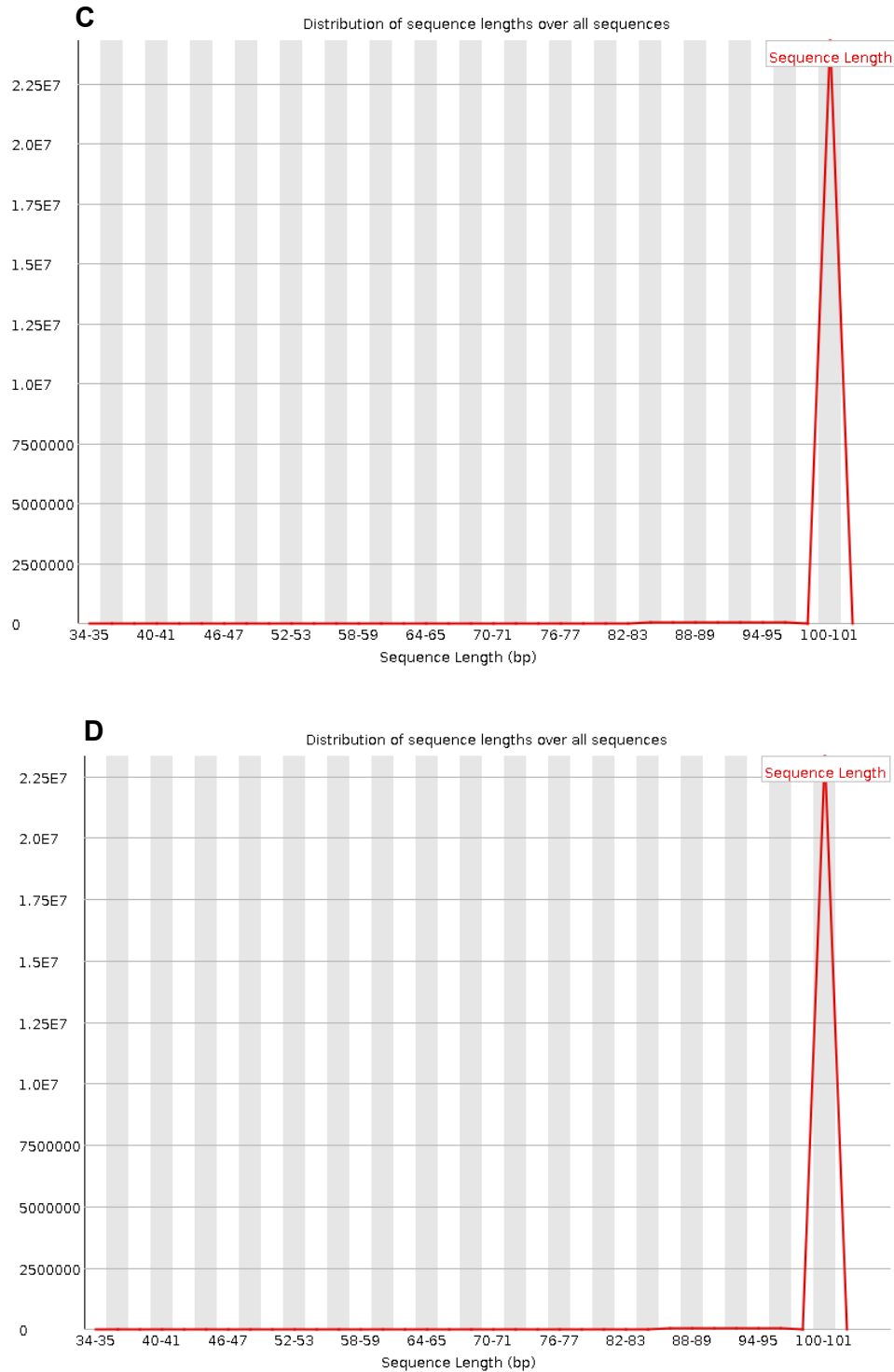


Figura Suplementaria 3.4. Distribución de la longitud de las secuencias (*reads*) del transcriptoma de frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro (SN) o suelo rojo (SR)) tras el *trimming* y filtrado de adaptadores, *reads* y bases de baja calidad. Cada transcriptoma tiene dos réplicas. A) SN1, B) SN2, C) SR1, D) SR2. El eje x representa la longitud de la secuencia y el eje y el número de secuencias. Las Figuras son representativas de las secuencias forward de las bibliotecas

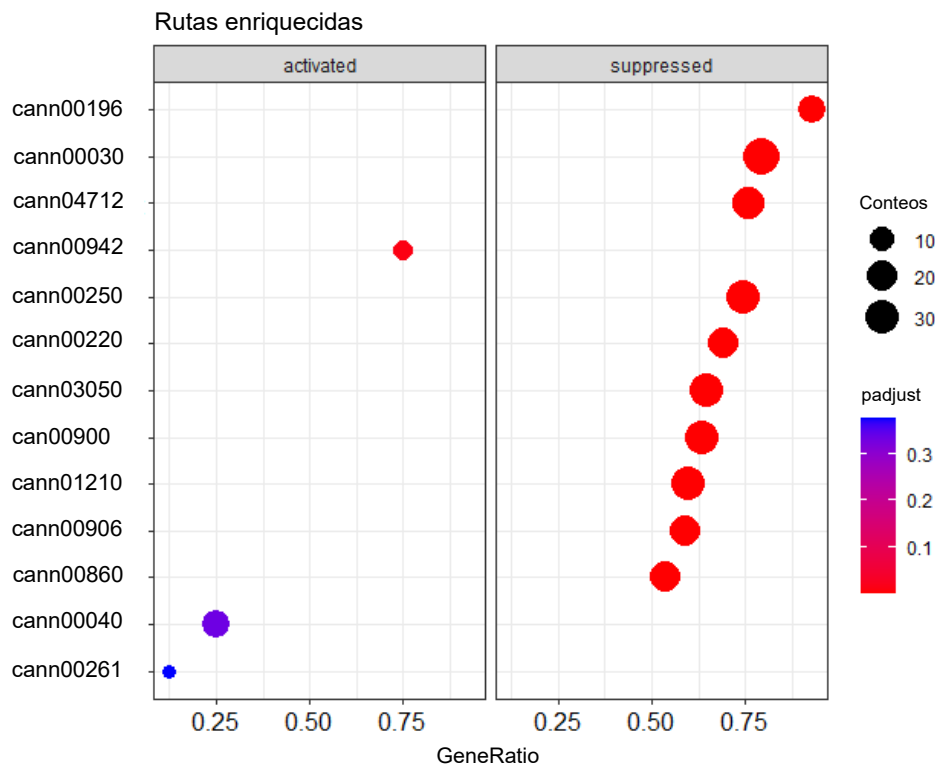
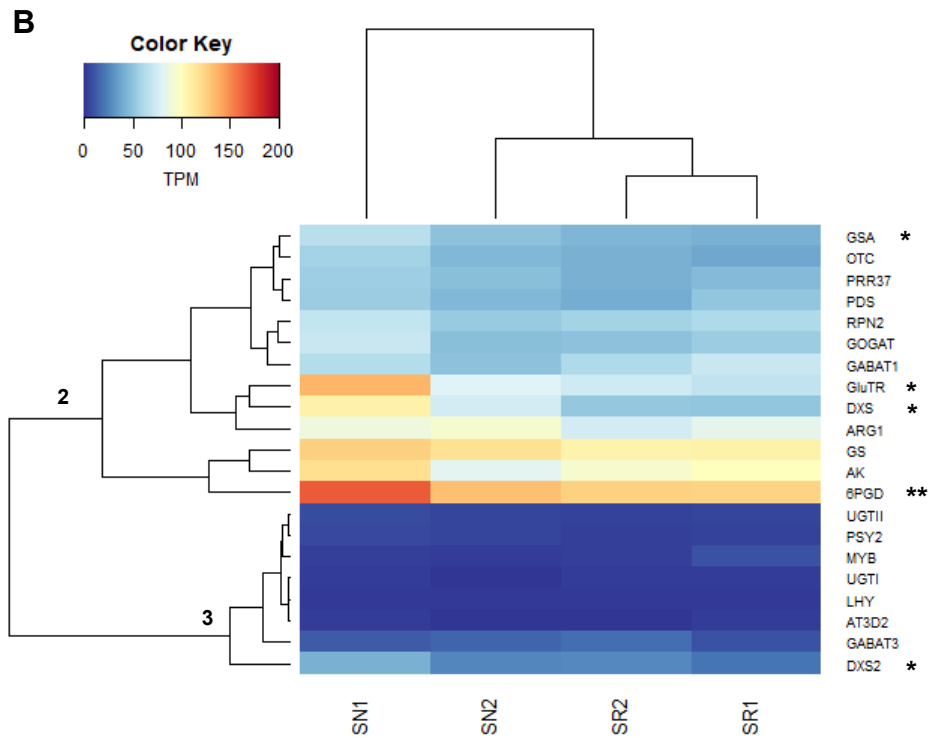
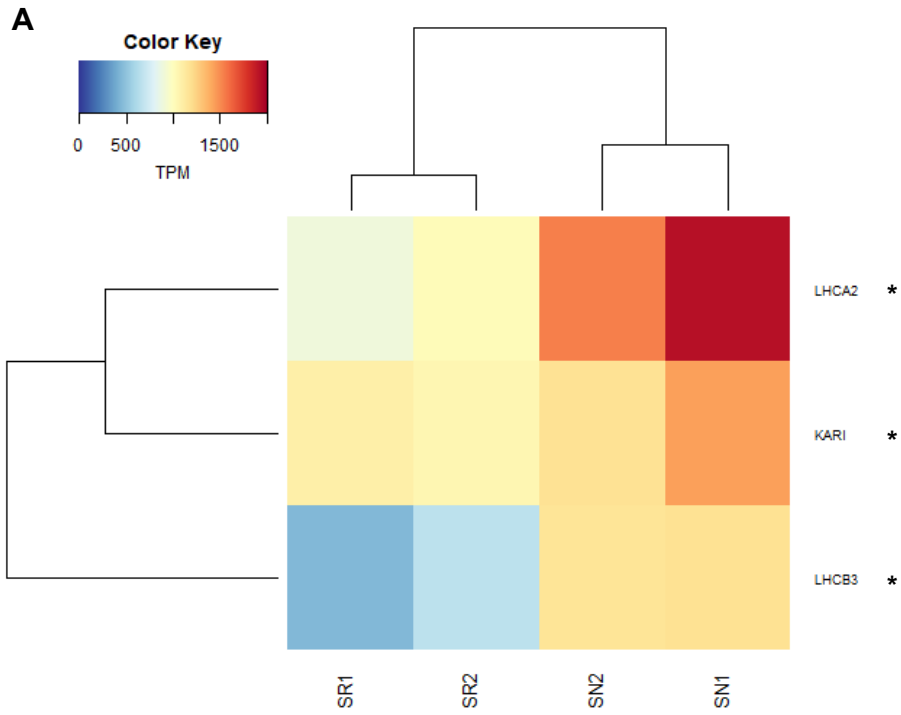


Figura Suplementaria 3.5. Análisis de enriquecimiento de genes (GSEA) por términos KEEG en la comparación de la expresión de los genes del transcriptoma en suelo rojo contra suelo negro. Las rutas de la izquierda (activadas) designan aquéllas sobrerrepresentadas en suelo rojo con respecto a suelo negro, y las del lado derecho (suprimidas) son aquéllas más enriquecidas en suelo negro con respecto a suelo rojo. El tamaño de los círculos representa el número de genes asociados con la ruta enriquecida correspondiente (conteos), y su color representa la confianza en la comparación (pajust), mientras que el eje x representa la proporción de genes asociados con una ruta metabólica en un suelo contra el número total de genes asociados esa vía (geneRatio). El eje y muestra las rutas metabólicas enriquecidas junto a su código en el KEEG. El GSEA se realizó por con ClusterProfiler v.4.4.4. cann00196: proteínas antena de fotosíntesis, cann00030: ruta de las pentosas fosfato, cann04712: ritmo circadiano, cann00942: biosíntesis de antocianinas, cann00250: metabolismo de alanina, aspartato y glutamato, cann00220: biosíntesis de arginina, cann03050: proteasoma, can00900: biosíntesis de terpenoides (backbone), cann01210: metabolismo del ácido 2-oxocarboxílico, cann00906: biosíntesis de carotenoides, cann00860: metabolismo de porfirina, cann00040: interconversión pentosa y glucuronato, cann00261: biosíntesis de monobactámicos



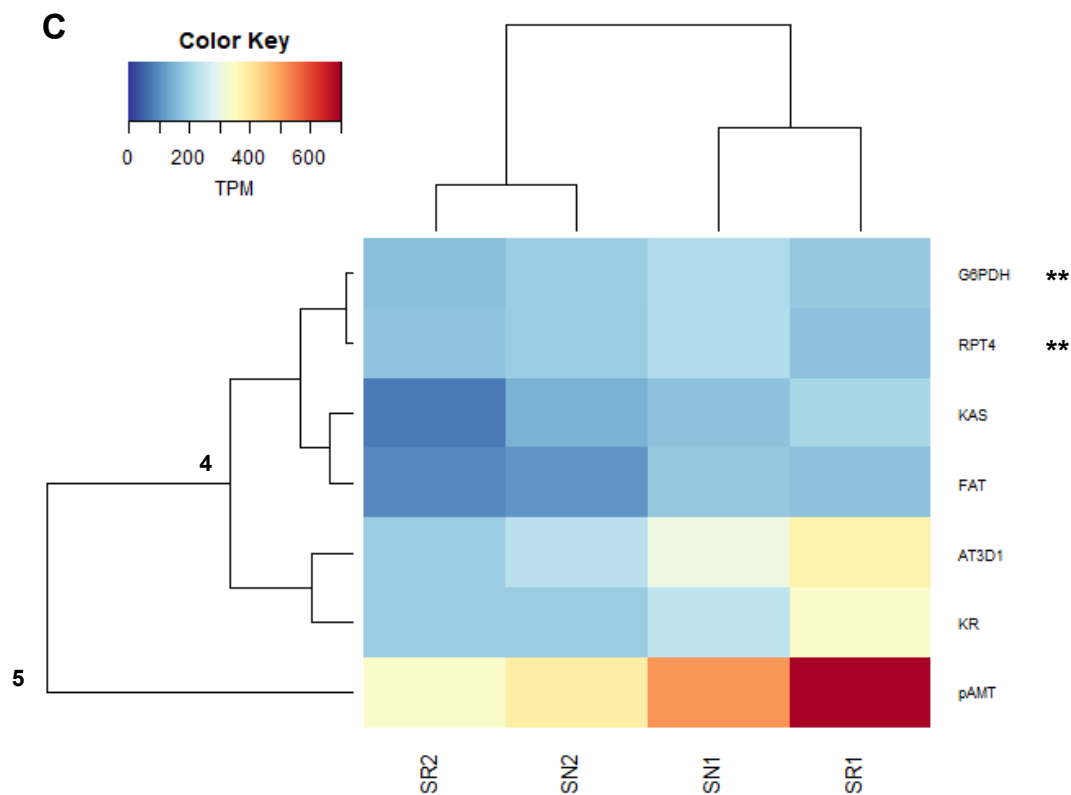


Figura Suplementaria 3.6. Mapas de calor de la expresión de genes, divididos en clados, de acuerdo al análisis de la Figura 3.6. A) Expresión de genes en el clado 1. B) Expresión de genes en los clados 2 y 3. C) Expresión de genes en los clados 4 y 5. Los valores de expresión están en transcritos por millón (TPM) correspondientes a las dos réplicas biológicas de frutos en suelo negro (SN1 y SN2) y suelo rojo (SR1 y SR2). La gráfica se realizó en R con la herramienta gplots. El código de los genes corresponde a la lista de la Tabla 3.3. *Genes diferencialmente expresados ($p_{adj} < 0.05$). ** Genes diferencialmente expresados ($p_{adj} < 0.1$) entre tipos de suelo de acuerdo al análisis de DESeq2

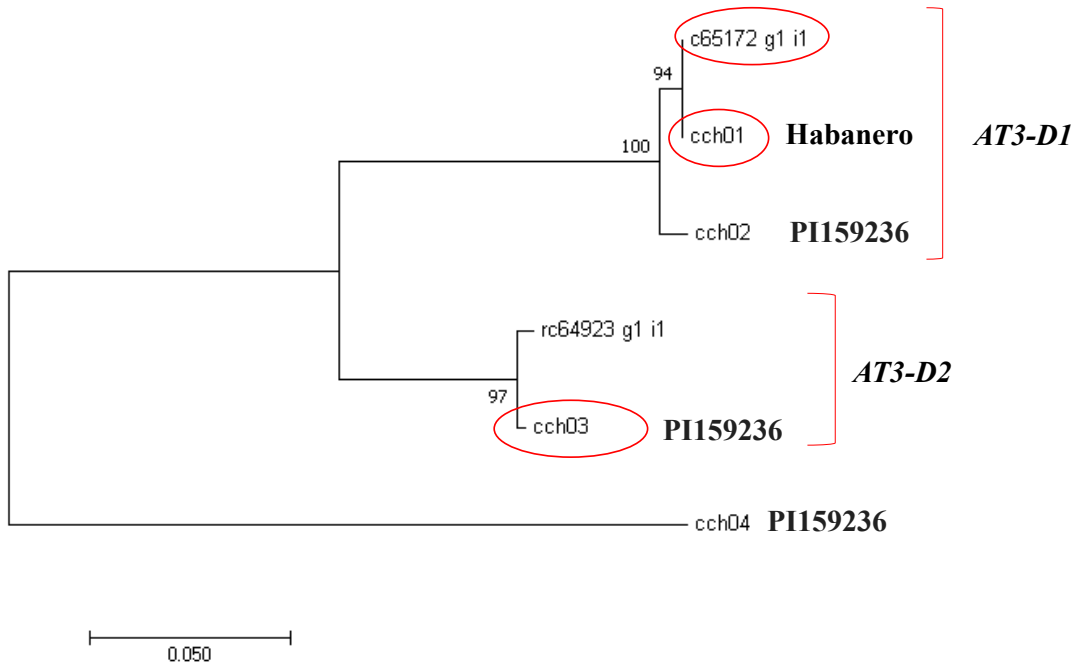
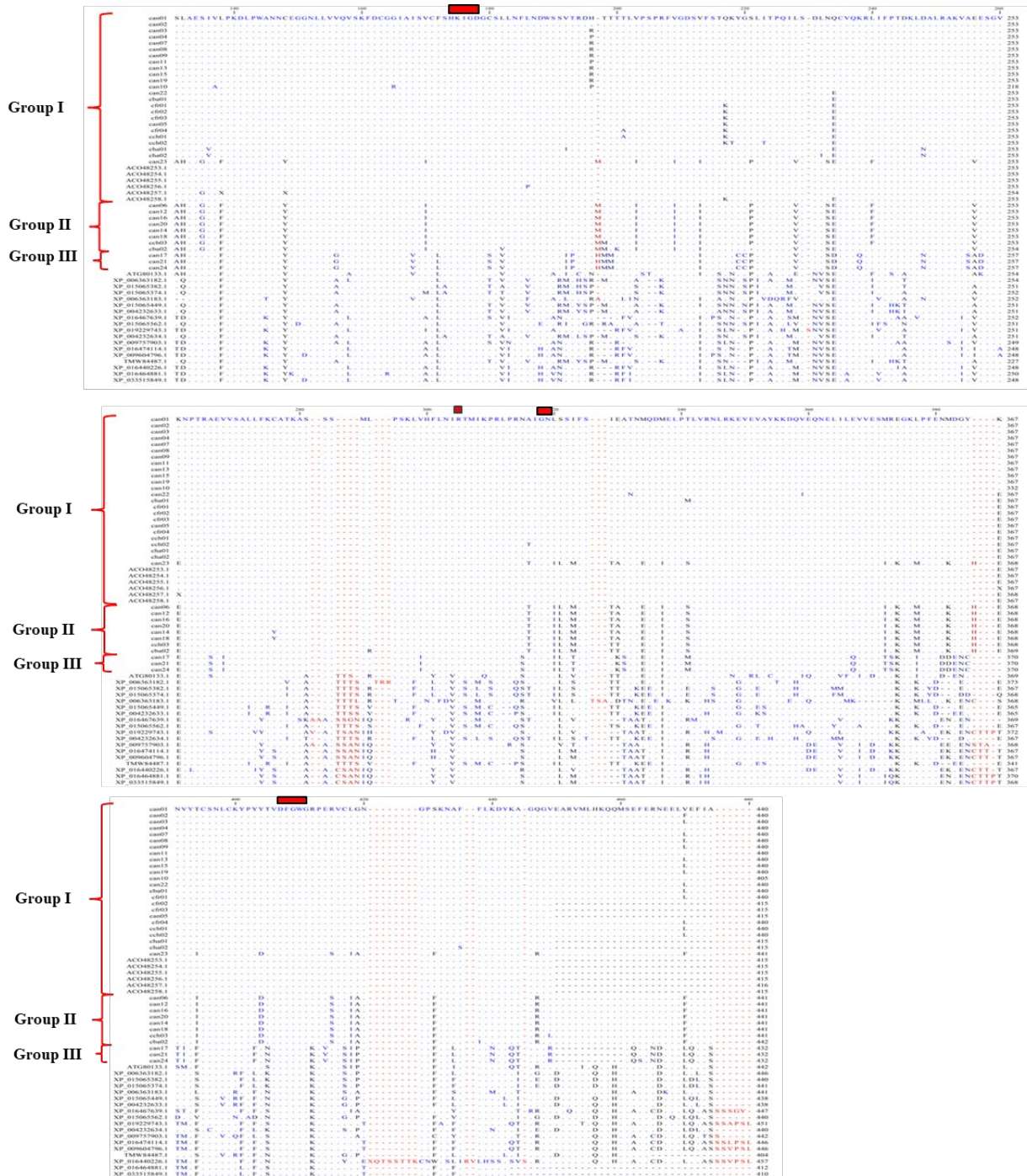
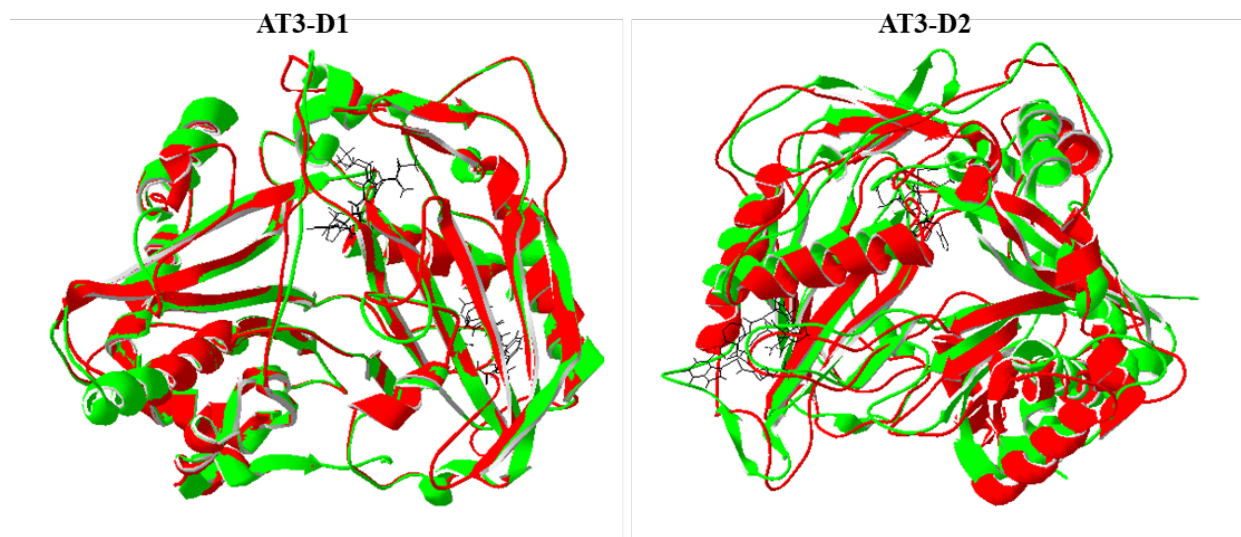


Figura Suplementaria 3.7. Análisis filogenético de los ortólogos c65172_g1_i1 y c64923_g1_i1 encontrados en la biblioteca de unigenes de los frutos de 25 DPA de plantas de chile habanero crecidas en suelo negro y suelo rojo. Las secuencias proteicas de otros ortólogos de *AT3-D1* y *AT3-D2* en *C. chinense* se usaron para agrupar las secuencias. Claves de las secuencias proteicas en el NCBI: cch01: AAV66309.1, cch02: PHU26684.1, cch03: PHU26685.1. La secuencia cch04 se encontró mediante un Blast en el genoma de *C. chinense* (PI159236) con secuencias de *AT3-D1* y *AT3-D2* como anzuelo (ccho4: CM008432.1:c149576995-149575438). El análisis se realizó haciendo un alineamiento múltiple por ClustalW y luego se construyó un árbol filogenético por Maximul likelihood con un bootstrap de 1000 usando el software Mega v.7.0.26



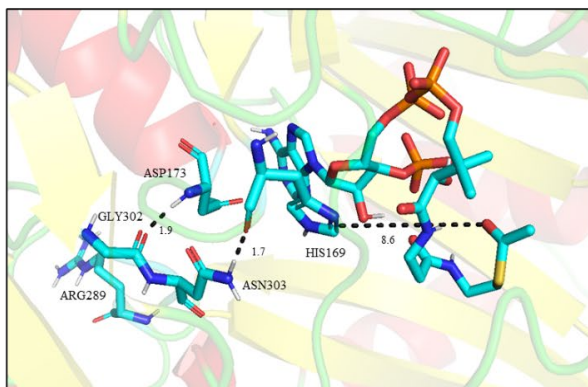
Supplementary Figure 4.1. Multiple alignment of 35 AT3 and homologue sequences. The conserved sequences HXXXD and DFGWG of the BAHD acyltransferases are signaled with a red bar. A GN sequence and a conserved arginine conserved in many acyltransferases are also signaled with a red bar. The more conserved residues are in blue and the less conserved are in red, while intermediate conservation are in dark tones. Groups I, II and III are the same groups formed in the phylogenetic and haplotype analysis in the main text. The alignment was made by ClustalW in the software Mega v.7.0.26. The figure was made with the software CLC sequence Viewer v.8.0.



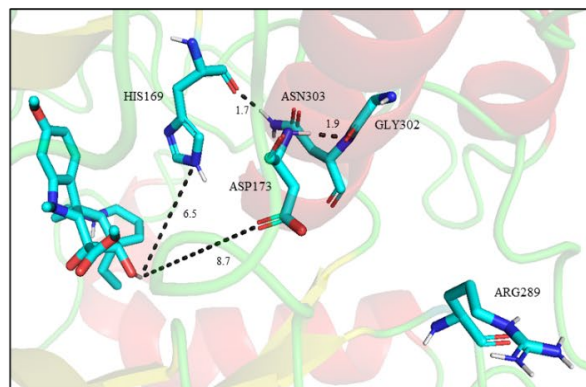
Supplementary Figure 4.2. Evaluation of AT3-D1 and AT3-D2 by comparison with templates. AT3-D1 (left) and AT3-D2 (right) alignment with the vinorine synthase and HCT, respectively. Vinorine synthase and HCT are in green color and the AT3 models in red color. The residues of the HXXXD and DFGWG motifs are in black sticks.

A

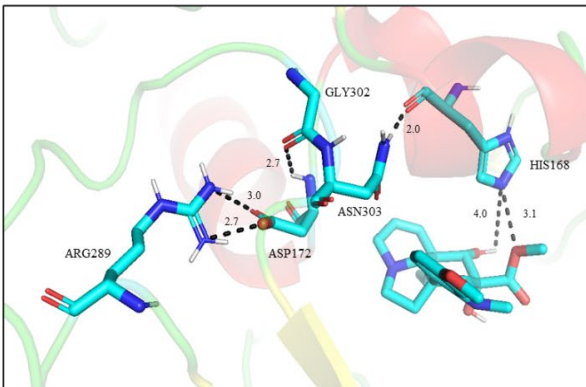
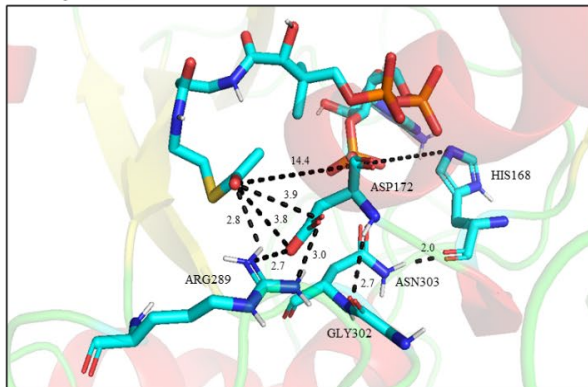
AT3-D1



B



AT3-D2



Supplementary Figure 4.3. Docking results of AT3 models with the DAT substrates. Acetyl-CoA (**A**) and 4-O-deacetylvindoline (**B**). Distances with nearest and catalytic residues are showed. AT3-D1 models (Up images), AT3-D2 models (Low images). Docking was made with Autodock Vina v.1.1.2. with AutodockTools v.1.5.6. Figures were made with PyMOL v2.3.0. Distances between atoms, in angstroms (Å), are signaled with a dotted line.