

Universidad Nacional de Misiones. Facultad de Ciencias Exactas, Químicas y Naturales. Doctorado en Ciencias Aplicadas

Doctorando
Lic. Sebastián Maximiliano ROSSI

Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. ST. HIL.

Tesis de Doctorado presentada para obtener el título de “Doctor en Ciencias Aplicadas”

“Este documento es resultado del financiamiento otorgado por el Estado Nacional, por lo tanto, queda sujeto al cumplimiento de la Ley N° 26.899”.

Director
Dr. Pedro Darío ZAPATA

Co-Director
Dr. Cristian FERRI

Posadas, 2023



Esta obra está licenciada bajo Licencia Creative Commons (CC) Atribución-NoComercial-Compartir Igual 4.0 Internacional <https://creativecommons.org/licenses/by-nc-sa/4.0/>

**“ANÁLISIS GENÓMICO COMPARATIVO, FILOGENÉTICO Y
DESCRIPTIVO DE LA VÍA BIOSINTÉTICA DE LA CAFEÍNA EN
Ilex paraguariensis A. ST. HIL.”**

**“COMPARATIVE GENOMIC, PHYLOGENETIC AND
DESCRIPTIVE ANALYSIS OF CAFFEINE BIOSYNTHESIS
PATHWAY IN *Ilex paraguariensis* A. ST. HIL.”**

Tesis Doctoral

LIC. SEBASTIÁN MAXIMILIANO ROSSI

AÑO 2023



Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. St.-Hilarie

Lic. Sebastián Maximiliano ROSSI

Tesis presentada a la Facultad de Ciencias Exactas, Químicas y Naturales
de la Universidad Nacional de Misiones para optar al grado académico
de **DOCTOR EN CIENCIAS APLICADAS**



POSADAS – MISIONES – ARGENTINA

AÑO 2023

*Tesis Doctoral
Sebastián Maximiliano Rossi*



Universidad Nacional de Misiones
Facultad de Ciencias Exactas, Químicas y Naturales



Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. St.-Hilarie

Lic. Sebastián Maximiliano ROSSI

DIRECTOR

Dr. Pedro D. ZAPATA

CO-DIRECTOR

Dr. Cristian FERRI



LUGAR DE TRABAJO

Laboratorio de Biotecnología Molecular

Instituto de Biotecnología Misiones “María Ebe Reca”

Facultad de Ciencias Exactas Químicas y Naturales (UNaM)

Ruta 12 Km 7 ½ – Campus Universitario – Posadas – Misiones

*Tesis Doctoral
Sebastián Maximiliano Rossi*

Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. St.-Hilarie

Lic. Sebastián Maximiliano ROSSI

COMISIÓN DE SUPERVISIÓN (Resolución Consejo Directivo N°

236/18) Dra. Sandra Pitta (INMIBO-UBA)

Dra. Marta Yajía (FCEQyN-UNaM)

Dra. Patricia Aguilera (IBS-INBIOMIS-UNaM-FCEQyN)

CARRERA DE DOCTORADO EN CIENCIAS APLICADAS

Proyecto de Carrera N° 10933/11

Con reconocimiento de la Comisión Nacional de Evaluación y

Acreditación Universitaria (CONEAU) N° 344/11

Tesis Doctoral
Sebastián Maximiliano Rossi

Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. St.-Hilarie

Lic. Sebastián Maximiliano ROSSI

COLABORACIÓN

Biological Sciences Department

Western Michigan University, EEUU

Dr. Todd J. BARKMAN



Dpto. de Química Biológica.

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires (UBA)

Dr. Adrián TURJANSKI





Análisis genómico comparativo, filogenético y descriptivo de la vía biosintética de la cafeína en *Ilex paraguariensis* A. St.-Hilarie

Lic. Sebastián Maximiliano ROSSI

DIRECTOR

Dr. Pedro D. ZAPATA

CO-DIRECTOR

Dr. Cristian FERRI

TRIBUNAL EXAMINADOR (Resolución Consejo Directivo N°376-20)

TITULARES

Dra. Sandra PITTA ÁLVAREZ (INMIBO-UBA)

Dr. Diego Alejandro SAMPIETRO (UNT)

Dra. Carolina URIBE CRUZ (UNaM)

SUPLENTES

Dra. Ana Isabel HONFI (IBS-FCEQyN-UNaM)

Dr. Julio Rubén DAVIÑA (IBS-FCEQyN-UNaM)

Dr. Emanuel GRASSI (IMiBio)

AGRADECIMIENTOS

Dedico este espacio para agradecer el enorme esfuerzo que conllevó la realización de este trabajo doctoral y las personas que directa o indirectamente estuvieron presentes y me ayudaron a culminar el doctorado desde lo profesional, personal y lo socio emocional, que me ayudaron a transitar una transformación desde la visión y experiencia de cada uno.

Es así que en primer lugar quiero dar gracias a mi familia, a mis padres Violeta y Pedro que siempre guiaron mis pasos personales y profesionales, con el respeto de mis decisiones y con la motivación de conocer el mundo, de explorar y explotar al máximo mis potenciales, por eso a los 16 años me emanciparon para que conozca Europa viajando solo, otras culturas, idiomas y personas, que me desenvuelva en un mundo que está en constante cambio para aprender y fortalecer mis propias habilidades cognitivas y emocionales por dos meses, con la absoluta confianza en mi madurez para desenvolverme solo, por ayudarme y motivarme a realizar mi tesis de licenciatura en la Universidad de Cornell, ciudad de Manhattan, Nueva York por un año y medio, a convencerme a mi mismo de no tener miedo y saber que para lograr el éxito hay que conseguir las cosas con convicción, actitud, trabajo y humildad, y continuar impulsando mis sueños estando cerca o lejos, pero siempre a mi lado, a mis hermanos Mariela, Martín y Andrea por acompañar cada paso en cada lugar que me tocó estar, estudiar y trabajar desde el respeto y apoyo incondicional, a mis sobrinos, tíos, primos y sobre todo a una persona que hasta en sus últimos momentos de vida me preguntaba sobre ciencias, era apasionada y siempre fue un placer charlar, mi abuela Verónica (1929-2020) hacedores del campo y de la yerba mate junto a mi abuelo Nicolás, campo que conocí desde mi infancia y aprendí el valor de la naturaleza, de lo sustentable, del cuidado de la vida, de la importancia para nuestro cerebro de la exploración, investigación y contacto con la naturaleza, pero por sobre todo de la unión y vínculo humano en un quehacer cotidiano desde el trabajo constante e incansable, la colaboración, la solidaridad y humildad, valores que construyen personas, sueños y futuros.

A mis amigos cercanos y lejanos que estuvieron siempre apoyando en las buenas y las malas, brindando y festejando con alegría cada paso, pero también cuando había que levantarse en las malas y continuar. Con un mensaje a la distancia, un consejo, una ayuda, una escucha, una caminata, un viaje, un mate y sobre todo sus tiempos, motivaciones y afectividad, por eso

gracias Rosina, Carolina, Horacio, Rebeca, Davina, Mariana, Gaby, Hernán, Alberto, Yanina, Fran, Sergio, Laura, Mechi, Emiliano, César, Cristina, Fer, Pablo, Silvia, Tamara, Maia, Dani, Seba y a todas las amistades que estuvieron en estos 5 años acompañándome.

A mi amigo Américo que ya no se encuentra más entre nosotros, en su memoria, con mucha alegría recordarlo como un profundo interesado por las ciencias, por horas de debate y charlas, por compartir toda mi carrera de grado y mitad de mi doctorado, por estar siempre presente para una pausa y disfrutar de la vida, de la amistad y la familia.

A mi director de tesis el Dr. Pedro Zapata por ser un mentor y amigo, por acompañar mis pasos desde mi carrera de grado como evaluador y ahora como guía científico y director doctoral. Por su calidez humana, sencillez, empatía, humildad y sabiduría. Por compartir momentos de debate, correcciones, cambios, reflexiones y acompañar e impulsar el trabajo en equipo con otros centros de investigación nacional e internacional, pensando siempre en alto y con solidez. Me enseñó la importancia de la gestión científica, de tener un espíritu crítico y superador siempre, de darme la confianza para la toma de decisiones y acompañarlas, por el respeto que me demuestra como persona y profesional, y por sobre todo por su amistad.

A mi co-director, Dr. Cristian Ferri por esa empatía eterna, bondad y alegría únicas y presentes en todo momento, por esa gigantesca motivación de lograr las metas desde la felicidad, desde lo humano, el buen trato, la eterna predisposición a ayudar y a escuchar. Por mostrar una sonrisa ante el problema y resolverlo. Me motivó a ver que cada vez faltaba menos y gracias a su buena onda y experiencia, pude clonar el primer gen de la vía metabólica después de un año de intentos fallidos, su optimismo, ánimos y alegría fue un motor potente para progresar y conseguir los resultados. ¡También se convirtió en un amigo!

A todo el equipo del Instituto de Biotecnología de Misiones, directivos y colegas de trabajo que me ayudaron con los cursos doctorales, trabajos en conjunto, compartimos diálogos, debates, materiales y recursos de investigación, consejos, mucha colaboración y también asados y reuniones after office. Gracias Gastón, March, Natalia, Marylin y Cecy de la "Big U".

Gracias al Dr. Martín Giorgio también del Instituto por guiar mis primeros pasos en este proyecto de investigación, por la puesta a punto de las técnicas, el acompañamiento, la docencia constante, por prestarme recursos, enseñarme metodologías y por acompañar de cerca este trabajo.

Tesis Doctoral
Sebastián Maximiliano Rossi

Agradezco infinitamente al Dr. Todd Barkman de la Western Michigan University de Estados Unidos quien fue mi tutor de tesis, colaborador directo de este proyecto de investigación y por brindarme la amistad suya y de su familia. Por haberme recibido tan gratamente en su

laboratorio junto a todo su equipo de trabajo, me ayudó a clonar los genes restantes de la vía y me brindó todos sus datos de investigación de años de esfuerzo de las vías metabólicas de cafeína en una enorme variedad de plantas. Le agradezco su enorme experiencia en genética evolutiva de plantas y bioquímica, por ser un extraordinario científico, amable, alegre, humano, trabajador incansable y por compartir momentos extracurriculares como hacer deportes, ir a recitales, compartir almuerzos y cenas, competencias de vinos mundiales y por hacerme un asado a pesar de ser vegetariano. Agradecido por la infinidad de emails en todos estos años trabajando en conjunto a la distancia, con muchos fracasos, frustraciones, pero con muchos éxitos al final. Por brindarme una oportunidad única con todo su equipo e integrarme a sus proyectos. Al Dr. Ruiqi Huang, extraordinario investigador de ese laboratorio que me acompañó en todo momento y me ayudó a utilizar herramientas de bioinformática, clonaciones, secuenciaciones, otro gran colaborador de este proyecto al brindarme el conocimiento de la filogenia de otras especies de plantas que evolucionaron con esta vía de biosíntesis.

Un enorme agradecimiento a todo el equipo del Dr. Adrián Turjanski del Dpto. de Química Biológica, laboratorio de bioinformática estructural de la UBA que el destino me llevó a conocerlos haciendo un curso doctoral en este mismo lugar y se convirtieron en los otros grandes colaboradores de este proyecto gracias a su trabajo se secuenciación completa del genoma de la yerba mate. Al Dr. Carlos Modenutti por brindarme las primeras secuencias para análisis, por los debates y sus tutoriales de ayuda. Un especial agradecimiento a mi mano derecha de trabajo en bioinformática del genoma y transcriptoma de la yerba y por incansables horas, días, semanas y meses de trabajo constante, el Dr. Federico Vignale con quienes trabajamos codo a codo en la escritura del paper con los datos del genoma ensamblado y la vía bioquímica de la cafeína, ayudé con el análisis de los metabolitos, evolución del genoma, abundancias relativas de genes como su incansable ayuda a la caracterización bioinformática de las enzimas de la vía metabólica, su modelado molecular y debate por plataformas virtuales en conjunto con el Dr, Barkman quien nos brindaba su experiencia en los aspectos evolutivos y metabólicos de las enzimas primarias y secundarias

*Tesis Doctoral
Sebastián Maximiliano Rossi*

de la vía de las diferentes especies de plantas que he utilizado a modo de genómica comparativa y filogenia. Hemos construido un lazo fuerte y sólido de trabajo entre WMU, UBA y UNaM gracias a este proyecto que nos conectó en los profesional y humano, y nos permitió publicar en conjunto desde el gran aporte de cada parte.

Agradezco al INTA Cerro Azul, especialmente al Ing. Mario Kryvenki por facilitarme el material de estudio de élite en varias ocasiones para la realización de la extracción del material genético

inicial, por su colaboración constante, predisposición y su carisma.

Por último, y no menos importante, agradecer de corazón a mi comisión de supervisión, a la Dra. Patricia Aguilera por motivarme a continuar desde su lado humano y empático, por apoyar, asesorar y guiar en cada seminario, por hacer las correcciones pertinentes y ayudarme en cada duda, le agradezco que, junto con el Dr. Mauro Grabiele, validaron la idea de pseudogenes gracias a sus investigaciones bioinformáticas de los datos del genoma y transcriptoma disponibles, y por sobre todo, la extrema buena onda de ambos.

A la Dra. Sandra Pitta y Dra. Marta Yajía por sus palabras de aliento en cada seminario, por sus tiempos de análisis y acompañamiento, por cada palabra superadora que motivaban a continuar y marcar el camino de cada avance. ¡Muchas gracias!

A todos quienes estuvieron, me escucharon, acompañaron, me entrevistaron y compartieron momentos desde lo laboral les digo gracias.

Gracias a todos quienes me permitieron crecer, aprender y superarme a mi mismo. Este trabajo es producto de una gran dinámica profesional, personal y emocional, y muestra la importancia del trabajo colaborativo y solidario en ciencias, del aporte diverso y multicultural, de los procesos de enseñanza y aprendizaje, de las convicciones, motivaciones y superaciones personales. ¡Por esto y mucho mas GRACIAS A TODOS!!

	AMP: Adenosina monofosfato
	BA: Benzoic Acid
	BAMT: Benzoic acid methyltransferase
	BLAST: Basic Local Alignment Search Tool
1X: 1-Metilxantina	
3X: 3-Metilxantina	BLASTn: Basic Local Alignment Search Tool for nucleotide
7X: 7-Metilxantina	
AAMT: Anthranilic acid methyltransferase	BLASTp: Basic Local Alignment Search Tool for protein
ADNc: Ácido Desoxirribonucleico complementario	BrEt: Bromuro de Etidio
AFLP: Amplified fragment length polymorphism	BTS1: Teobromina sintasa de <i>Theobroma</i> Ca: <i>Coffea arabica</i>

CaDXMT: *Coffea arabica* 3, 7-

dimethylxanthine methyltransferase

Tesis Doctoral

Sebastián Maximiliano Rossi

EDTA: Ácido

Etilendiaminotetraacético **EST:**

Expressed Sequence Tag

LISTA DE ABREVIATURAS

FAMT: Farnesoic acid methyltransferase

GAMT: Gibberellic acid

methyltransferase **GMP:** Guanosina
monofosfato

HMM: Hidden Markov Model

HPLC: Cromatografía líquida de alto
rendimiento

JMT: Jasmonic acid methyltransferase

IAMT: Indole-3-acetic acid
methyltransferase

IMP: Inosine monophosphate

INTA: Instituto Nacional de Tecnología
Agropecuaria

INYM: Instituto Nacional de la Yerba

Mate **Ip:** *Ilex paraguariensis*

IpCS1: Cafeína sintasa 1 de *Ilex*
paraguariensis

IpCS2: Cafeína sintasa 2 de *Ilex*
paraguariensis

IpCS3: Cafeína sintasa 3 de *Ilex*
paraguariensis

IPTG: Isopropil-B-D-

Tiogalactopiranósido **ITS:** Internal
Transcribed Spacer

JTT: Jones-Taylor-Thornton model

CaMXMT: *Coffea arabica* 7-

methylxanthine methyltransferase

CaXMT: *Coffea arabica*

Xanthine/xanthosine

methyltransferase **Cb:** *Clarkia breweri*

Cc: *Coffea canephora*

CcDXMT: *Coffea canephora* 3, 7-
dimethylxanthine methyltransferase

CCMT: Cinnamate/p-coumarate
methyltransferase

CcMTL: *Coffea canephora*
Methyltransferase-like gene

CcNMT3: *Coffea canephora* N
metiltransferasa 3

CF: Cafeína

Cis: *Citrus sinensis*

CisXMT1: Citrus XMT1

CS: Cafeína sintasa

CTAB: Cetyl Trimethyl Ammonium
Bromide

DXMT: 3, 7-dimethylxanthine
methyltransferase

LAMT: Loganic acid methyltransferase	
LC-MS: Cromatografía líquida - Espectrometría de masas en tándem	RT-PCR: Reverse Transcription – Polymerase Chain Reaction
MAFFT: Multiple Alignment using Fast Fourier Transform	SA: Salicylic Acid
MBMT: p-methoxybenzoic acid methyltransferase	SABATH: SAMT, BAMT, theobromine synthase
MTs: Methyltransferases	SAM: S-adenosyl-L-methionine SAMT: salicylic acid methyltransferase
MXMT: 7-methylxanthine methyltransferase	SDS-PAGE: geles de poliacrilamida desnaturalizantes
NAMT: Nicotinic acid methyltransferase	SNP: Single Nucleotide Polymorphism TB: Teobromina
NCBI: National Center for Biotechnology Information	Tc: Theobroma cacao
NGS: Next Generation Sequencing ORF: Open Reading Frame	
Pc: <i>Paullinia cupana</i>	
PcCS: <i>Paullinia cupana</i> cafeína sintasa	TCS2: Cafeína sintasa 2 de <i>Camellia</i>
PcCS1: <i>Paullinia cupana</i> cafeína sintasa 1 PcCS2: <i>Paullinia cupana</i> cafeína sintasa 2 PM: Peso Molecular	TCS1: Cafeína Sintasa 1 de <i>Camellia</i>
PROMATEAR: Proyecto de Secuenciación del Genoma de Yerba Mate	Tm: Temperature of melting
PX: Paraxantina	TP: Teofilina
RAPDs: Random Amplification of	WGD: Whole-genome Duplication X: Xantina
	XR: Xantosina
	XMP: xanthosine monophosphate
	XMT: Xanthine/xanthosine methyltransferase
	XR: Xantosina

Tesis Doctoral
Sebastián Maximiliano Rossi

Tesis Doctoral
Sebastián Maximiliano Rossi

FIGURA 1: La cafeína ha evolucionado de manera convergente en cinco especies de plantas con flores utilizando diferentes combinaciones de genes y vías. (A) Las relaciones filogenéticas entre órdenes de Rósidas y Astéridas muestran múltiples orígenes de la biosíntesis de cafeína. Los linajes de color verde lima rastrean el antiguo linaje CS de enzimas que ha sido reclutadas de forma independiente para su uso en tejidos que acumulan cafeína en *Theobroma*, *Paullinia* y *Camellia*. Linajes turquesa trazan el antiguo

linaje XMT que fue reclutado independientemente en *Citrus* y *Coffea*. (B y C) *Theobroma* 16 y *Paullinia* han convergido en vías biosintéticas similares catalizadas por enzimas de tipo CS. (D) *Citrus* ha evolucionado una vía diferente catalizada por enzimas de tipo XMT, a pesar de su estrecha relación con *Paullinia*. (E y F) *Camelia* y *Coffea* catalizan la misma vía utilizando diferentes enzimas. Las vías bioquímicas propuestas se basan en actividades enzimáticas relativas mostradas por gráficos de barras correspondientes que indican actividades relativas medias (de 0 a 1) con ocho sustratos de alcaloides de xantina. CisXMT1 y TCS1 catalizan más de una reacción en las vías propuestas. XMT y CS se han duplicado recientemente e independientemente en cada uno de los cinco linajes. **Fuente:** Los datos fueron tomados con permiso del autor del trabajo publicado “*Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes*, Huang et al, 2016, colaborador directo del presente trabajo doctoral.

FIGURA 2: Biosíntesis de cafeína en *Coffea arabica*. Ésta vía de biosíntesis se lleva a cabo por enzima de la familia SABATH de tipo XMT. El primer paso de formación de 7- metilxantina está dado por CaXMT a partir de Xantosina. Para formar teobromina, otra

18

enzima la CaMXMT cataliza dicha formación a partir de 7-metilxantina y CaDXMT es la responsable del último paso de la vía para formar cafeína a partir de teobromina. **Fuente:** Adaptado de Huang, R. “*Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants*” (2017).

FIGURA 3: Biosíntesis de cafeína en *Camellia sinensis*. La enzima TCS2 cataliza el primer paso de la vía convirtiendo xantosina a 7-metilxantina, mientras que la enzima

20

TCS1 tiene una acción dual sobre 7-metilxantina y teobromina para formar cafeína. **Fuente:** Adaptado de Huang, R. “*Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants*” (2017).

FIGURA 4: Biosíntesis de cafeína en *Paullinia cupana*. La síntesis de cafeína comienza a partir de enzimas codificadas por genes tipo CS caracterizadas a partir semillas de *Paullinia cupana*. La vía comienza con la xantina, y los intermediarios son 3-metilxantina 21

y teobromina, diferentes de las encontradas en vías en *Camellia*, *Coffea* y *Citrus*. Estas tres enzimas distintas PcCS1=PaulliniaCS1, PcCS2=PaulliniaCS2 y PcCS son responsables de cada paso de metilación en la ruta. **Fuente:** Adaptado de Huang, R. “*Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants*” (2017).

FIGURA 5: Biosíntesis de cafeína en *Citrus sinensis*. La cafeína se sintetiza utilizando genes del linaje XMT a través de dos nuevas vías en las flores de cítricos. Las vías

22

comienzan con la xantina, en lugar de la xantosina, y los intermedios son 1-metilxantina, 3-metilxantina y teofilina, completamente diferentes de la vía propuestas en *Camellia* y *Coffea*. Dos enzimas, CisXMT1=CitrusXMT1 y CisXMT2= CitrusXMT2 son responsables de

Tesis Doctoral
Sebastián Maximiliano Rossi

los tres pasos de metilación y CitrusXMT1 es una enzima multifuncional que cataliza los dos primeros pasos de metilación en cada vía. **Fuente:** Adaptado de Huang, R. “*Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants*” (2017).

FIGURA 6: Estructura 3D de la cafeína sintasa en *Ilex paraguariensis*. (CS). Empleando la estructura resuelta de cristalografía de rayos X de CS de *Coffea arabica* como molde

(b), la estructura 3D de la yerba mate CS fue predicha por el algoritmo del SWISS-model

24

(a). Un modelo de cinta de CS de yerba (c) y CS de café (e) sugiere una alta conservación de la estructura secundaria cuando se superponen (d). Se presenta una reconstrucción de un modelo de malla de yerba CS (f) y se compara con el café EM (h), que muestra una gran similitud de estructura cuaternaria (g). **Fuente:** Debat *et al.* 2014.

FIGURA 7: Cladograma en el cual se expresan las interrelaciones entre órdenes y familias de Angiospermas aportadas por el análisis de Jackknife con frecuencias **31**

superiores a 50% (APG II, 2003 De acuerdo al APG II (2003) y a Stevens (2001) los principales grupos, formales y no formales, de las Angiospermas son los que siguen a continuación. En el recuadro rojo se indica el orden a la cual pertenece *Ilex paraguariensis* dentro del grupo de las euastéridas II.

FIGURA 8: Representación esquemática de la hipótesis de Horowitz sobre el origen y **36** la evolución de las vías metabólicas (modificado de Fondi *et al.* 2009). Las secuencias de enzimas que hacen transformaciones sucesivas a una sustancia A, para dar lugar a otra, B, y luego a un producto final C, se habrían originado hacia atrás.

FIGURA 9: Se muestra un modelo esquemático de tres pasos de la hipótesis del modelo Patchwork: **(a)** una enzima ancestral E0 dotada de baja especificidad de sustrato puede unirse a tres sustratos (S1, S2 y S3) y catalizar tres reacciones diferentes, pero similares. ; **(b)** una duplicación del gen que codifica E0 y la posterior divergencia de una

38

de las dos copias conduce a la aparición de la enzima E2 con una especificidad aumentada y reducida; y **(c)** un evento adicional de duplicación de genes, seguido de divergencia evolutiva, conduce a E3. De esta manera, la enzima ancestral E0 que pertenece a una ruta metabólica dada es "reclutada" para servir a otras vías novedosas. **Fuente:** modificado de Fani, 2012.

FIGURA 10: Comparación esquemática de las tres hipótesis sobre el origen y la evolución de las vías metabólicas en plantas. La hipótesis retrógrada (a) es opuesta a la hipótesis acumulativa (b) en el orden (tiempo) en el que se ensamblaron los pasos. Sin **40** embargo, ambos afirman que el camino se ensambló de manera gradual. La hipótesis en retazos o modelo patchwork (c) se centra en el reclutamiento de enzimas que catalizaron reacciones análogas en una nueva vía al hacer uso de sus actividades promiscuas. La duplicación de genes y la divergencia funcional de estas enzimas proporcionaron la nueva vía con enzimas recientemente especializadas.

FIGURA 11: Árbol filogenético que indica de dónde se derivan las secuencias RSA, así

42

como la diferencia entre la comparación horizontal y vertical. (Dr. Lewis Pirenne, 2016)

FIGURA 12: Árbol filogenético de órdenes de angiospermas correspondientes a **44** Astéridas y Rósidas. Las líneas de colores muestran que al menos 5 veces ha evolucionado la vía metabólica de la cafeína en determinados grupos de plantas de eudicotas.

FIGURA 13: Representación esquemática de la formación de genes parálogos (A) y de **46** familias de genes parálogos (B). **Fuente:** Modificado de Fani, 2004.

Tesis Doctoral
Sebastián Maximiliano Rossi

FIGURA 14: Elongación génica. La duplicación de un gen ancestral y la posterior fusión **47** de los dos homólogos para producir una proteína más larga. **Fuente:** modificado de Fani, 2004.

49

FIGURA 15: Estructuras de alcaloides de purina presente en plantas que constituyen las metilxantinas

FIGURA 16: La vía biosintética de la cafeína en las especies de plantas estudiadas que son capaces de producirla, tienen al menos 12 potenciales vías de producción. La única vía bien caracterizada de plantas se muestra con flechas negras e involucra una metilación secuencial de xantosina en el N-7 del anillo heterocíclico como así también a nivel del N-3 de la 7-metilxantina y N-1 de la teobromina. Cada paso de metilación es

llevado a cabo por distintas metiltransferasas de alcaloides de xantina en *Coffea*. En 53
contraparte, *Camellia* emplea una enzima relacionada distantemente, la cafeína sintasa,
TCS1 para el segundo y tercer paso de metilación, mientras que la enzima que cataliza el
primer paso no se encuentra caracterizada. Otras vías bioquímicas potenciales de producción
de cafeína se muestran en flechas discontinuas, pero las enzimas especializadas para dichas
conversiones son desconocidas. El clivaje de ribosa por 7- metilxantosina no se muestra, pero
debería ocurrir de forma concomitante con la metilación de N-7 de xantosina. CF, cafeína;
PX, paraxantina; TB, teobromina; TP, teofilina; X, xantina; 1X, 1-metilxantina; 3X, 3-
metilxantina; 7X, 7-metilxantina; XR, xantosina (modificado de *Huang et al. 2016*).

FIGURA 17: Formación de xantosina para la biosíntesis de cafeína a partir de nucleótidos de purina y SAM. La
xantosina se produce a través de al menos cuatro rutas: 56

a partir de IMP que se origina en la síntesis de novo de purina (ruta de novo), del pool de
nucleótidos de adenina celular (ruta AMP), de adenosina liberada del ciclo SAM (ruta SAM)
y del pool de nucleótidos de guanina (ruta GMP). **Fuente:** "Distribution, Biosynthesis and
Catabolism of Methylxanthines in Plants". Hiroshi Ashihara, 2010.

FIGURA 18: Comparación de las secuencias de aminoácidos de la cafeína sintasas y sus
enzimas relacionadas. Alineamiento de las secuencias de aminoácidos para TCS1 y TCS2 de
té, ICS1 e ICS2 de *Camellia irrawadiensis*, PCS1 y PCS2 de *Camellia ptilophylla*, BTS1 de cacao
y se indica CmXRS1, CTS1 y CCS1 del café. Los cuadros sombreados representan residuos de
aminoácidos conservados y los guiones representan espacios o "gaps" que se han insertado
para una alineación óptima. Los motivos de unión SAM

59

propuestos (A, B' y C) y la "región YFFF" conservada se muestran por cajas abiertas (Mizuno
et al. 2003). Los asteriscos indican residuos de tirosina (Y) o fenilalanina (F) en la región. Los
aminoácidos nominados en la unión del sustrato se indican mediante círculos cerrados, y los
residuos adicionales del sitio activo se indican con puntas de flecha (Zubieta *et al.* 2003). Las
fuentes de las secuencias son las siguientes: TCS1, AB031280 (Kato *et al.* 2000); TCS2,
AB031281; BTS1, AB096699; PCS1, AB207817; PCS2, AB207818; ICS1, AB056108; ICS2,
AB207816 (Yoneyama *et al.* 2006); CmXRS1, AB034699 (Mizuno *et al.* 2003); CTS1, AB034700
(Mizuno *et al.* 2001); CCS1, AB086414 (Mizuno *et al.* 2003). (Adaptado de Yoneyama *et al.*
2006).

61

FIGURA 19: Representación molecular de la cafeína

FIGURA 20: Metabolismo de la cafeína por desmetilación del isoenzima del citocromo 63
P-450 (CYP) hepático, subfamilia 1A, gen 2 (CYP1A2). Aznar, 2011.

FIGURA 21: Análisis microsintético de tres especies *Theobroma cacao*, *Citrus sinensis* 67
y *Coffea canephora* de bloques génicos mostrando en colores los HSP (High- Scoring segment
Pair) o par de segmentos de alta puntuación, es un subsegmento de un par de secuencias,
ya sea de nucleótidos o de aminoácidos, que comparten un alto nivel de

Tesis Doctoral
Sebastián Maximiliano Rossi

similitud. El nivel de similitud entre las secuencias depende de la configuración del
algoritmo de alineación local. **Fuente:** NCBI web site

73

FIGURA 22: Estrategias de amplificación del gen IpCS1 para maximizar las posibilidades
de obtención de bandas con el tamaño deseado.

79

FIGURA 23: Plásmido pTrcHis-TOPO

FIGURA 24: cDNA del producto RT-PCR. Se llevó a cabo mediante dos extracciones de
ARN diferentes utilizando el mismo protocolo (ARN1 y ARN2). Para cada ARN se procedió

88

a realizar una PCR con enzimas taq polimerasa distintas (pFu y GoTaq) para maximizar el
éxito de la obtención del producto. Los cebadores F y R utilizados en cada caso fueron los
mismos. L= Ladder 1kb invitrogen.

FIGURA 25: Sección del cromatograma mostrando algunos SNPs en la 89
secuenciación. **A:** Secuencia Forward. Se denota dos picos superpuestos en el
cromatograma, en este caso se superponen una A y G en la posición 104. **B:** se observa otro
SNP en la posición 570 de una G y C.

FIGURA 26: CLUSTAL format alignment por MAFFT. Alineamiento de secuencias 90
aminoácidas de los constructos de la colonia A y Z. Se muestran las diferencias de
aminoácidos en las cajas sombreadas. El porcentaje de identidad es del 94,25%.

FIGURA 27: a) Paso de la vía que involucra la actividad de la enzima de la colonia Z catalizando la formación de
teobromina (TB) a partir de 3-Metilxantina (3X). **b) Ensayo 91**
enzimático utilizando la enzima de la colonia Z. Se muestra alta actividad catalítica para el
sustrato 3-Metilxantina indicando interconversión a Teobromina. Sustratos utilizados: 1)
Xantina, 2) Xanthosina, 3) 1-Metilxantina, 4) 3-Metilxantina, 5) 7-Metilxantina, 6) Teofilina,
7) Teobromina, 8) Paraxantina.

FIGURA 28: Alineamiento de la secuencia aminoácida del genoma 3 (*Query*) con la
secuencia de la colonia Z (*Sbjct*), mostrando en el recuadro rojo una inserción de 9

92

aminoácidos que según el análisis genómico posterior fue descartado debido a caer en una
zona intrónica. Por ende, la proteína utilizada para los ensayos enzimáticos presentaba los
365 aminoácidos de la secuencia consenso.

FIGURA 29: Alineamiento secuencia del transcriptoma 3 (*Query*) con colonia Z (*Sbjct*). 93
Se puede observar respecto a la figura 30 que no se encuentra la inserción anteriormente
mencionada en la secuencia del transcriptoma.

FIGURA 30: Alineamiento de la secuencia aminoácida del genoma 4 (*Query*) y el de

93

la colonia A (*Sbjct*). El recuadro rojo señala una inserción de 9 aminoácidos del genoma
diferente a las de la figura 30.

FIGURA 31: Alineamiento de la secuencia aminoácida del genoma 4 (*Query*) con la

94

secuencia traducida del transcriptoma 4 (*Sbjct*).

FIGURA 32: Comparación de secuencias aminoácidas entre el transcrito #4 (*Query*)

94

y colony A (*Sbjct*).

FIGURA 33: Alineamiento de la secuencia del transcriptome 2 (*Sbjct*) que es el

95

transcrito de la secuencia genómica 2 (*Query*).

Tesis Doctoral

Sebastián Maximiliano Rossi

FIGURA 34: Alineamiento y comparación de las
secuencias seleccionadas a sintetizar. "*Query*"
corresponde a la secuencia del genoma 2 y "*Sbjct*" a la
secuencia del genoma 4, como candidatas a la vía de
biosíntesis de la cafeína.

FIGURA 35: Estas secuencias fueron obtenidas
utilizando la herramienta de *Hidden Markov Model* con
los perfiles de N-metiltransferasas involucrados en la
biosíntesis de cafeína, comparando los datos del
transcriptoma y el reciente genoma de la YM
secuenciado. Dentro del genoma de *I. paraguariensis*
todos estos clústeres de genes se encuentran juntos en el
mismo scaffold. En el recuadro rojo se comparan de la
primera secuencia clonada y secuenciada "Colony A"
como candidata CS de la vía de biosíntesis con los datos

del transcriptoma y genoma con mayor homología de
secuencia, genome_4 y transcriptome_4. Genome_4 fue
sintetizada. En el recuadro verde, se puede observar la
otra secuencia primariamente clonada y secuenciada
"Colony Z" que se demostró presentar actividad catalítica
para convertir TB a partir de 3X comparando con las
secuencias de mayor homología genome_3 y
transcriptome_3. En el recuadro azul, se encuentra la
secuencia candidata la vía de biosíntesis que fue
sintetizada, clonada y expresada (genome_2) por
presentar una longitud típica de enzimas CS SABATH a
diferencia de las otras secuencias aquí mostradas pero
incompletas a nivel de los extremos o secuencias
internas.

FIGURA 36: Se muestran alineamientos de secuencias
del genoma 1, 2, 3 y 4 con secuencias del transcriptoma 2,

3 y 4 comparando con las secuencias de colonia A y Z en regiones que se relacionan a la preferencia por el sustrato. **A)** región 20-30 del alineamiento mostrando diferencias de aminoácidos entre las secuencias 2 y 4 del genoma y la secuencia 4 del transcriptoma y **B)** región 320-330 mostrando diferencias de las secuencias del genoma 4 y colonia A con una serina (S) en el sitio de una cisteína (C), como así también una fenilalanina (F) por una tirosina (Y) al igual que la secuencia 2 del genoma en éste último aminoácido.

FIGURA 37: Análisis bioinformático de las secuencias genómicas G12, G23 y G24 candidatas a cafeína sintetas de acuerdo al árbol filogenético realizado. Se muestran las comparaciones de las secuencias originales (V1) con las expresadas predichas (V2) mostrando segmentos faltantes o sin homología de secuencias definidos como pseudogenes.

FIGURA 38: Cromatogramas de HPLC midiendo las lecturas de la actividad catalítica de las enzimas identificadas de participar en la vía de biosíntesis de la cafeína. **A.** Patrones de medición standart de cada sustrato. **B.** IpCS1 mostrando actividad catalítica con el sustrato 3-metilxantina. **C.** IpCS2 reacciona con teobromina luego de ser suministrado con 3X. **D.** IpCS3 al ser suministrado con teobromina muestra un pico de reacción con cafeína.

FIGURA 39: Las enzimas SABATH han evolucionado para catalizar la biosíntesis de cafeína en *Ilex paraguariensis*. **A.** Vía biosintética propuesta para la cafeína en la yerba mate. **B.** Actividades enzimáticas relativas de tres enzimas SABATH de tipo CS codificadas con ocho sustratos de alcaloides de xantina. **C.** Trazas de HPLC que muestran productos formados por tres enzimas codificadas de tipo CS. X = xantina, XR = xantosina, 1X = 1-metilxantina, 3X = 3-metilxantina, 7X = 7-metilxantina, TP = teofilina, TB = teobromina, PX = paraxantina

FIGURA 40: La organización de genes en el scaffold 93 revela la presencia de genes de acetiltransferasa y genes de la familia de proteínas transportadoras de oligopéptido dependientes de protones que comprenden a los genes de la vía biosintética de cafeína.

96

Tesis Doctoral
Sebastián Maximiliano Rossi

FIGURA 41: Vía metabólica completa de biosíntesis de cafeína en *Camellia sinensis*. La primera vía de comparación para obtener los primeros resultados en *Ilex paraguariensis*. La enzima TCS1 tiene una acción dual ya que cataliza dos pasos de la vía, de 7-metilxantina a teobromina y de teobromina a cafeína. La última enzima caracterizada de la vía, TCS2 puede catalizar la formación de 7-metilxantina a partir de xantosina. ácido logánico/farnesoico MT, SAMT = ácido salicílico MT, XMT = alcaloide MT de xantina utilizado para la biosíntesis de cafeína en *Coffea* y *Citrus*, BAMT = ácido benzoico MT, JMT = ácido jasmónico MT, CS = cafeína sintasa en *Ilex*, *Theobroma*, *Camellia* y *Paullinia*.

FIGURA 42. El genoma de *Ilex paraguariensis* codifica varias enzimas SABATH distintas que son miembros de varios linajes funcionalmente caracterizados. Los puntos indican genes de longitud completa en *I. paraguariensis* y los clados a los que pertenecen están etiquetados de acuerdo con las funciones publicadas. GAMT = Giberelina MT, IAMT = ácido indol-3-acético MT, LAMT/FAMT =

FIGURA 43: Análisis filogenético por el método de máxima similitud (1,000 repeticiones bootstrap) de Secuencias de aminoácidos de N-metiltransferasas de *I. paraguariensis* con secuencias representativas de N-metiltransferasas del género *Triticum*, *Glycine*, *Populus*, *Hevea*, *Jatropha*, *Citrus*, *Gossypium*, *Coffea*, *Chenopodium*, *Durio*, *Theobroma*, *Eucalyptus*, *Punica* y *Camelia*. Secuencias de aminoácidos de la N metiltransferasa de *Zea mays*, *Oryza sativa* y *Sorghum bicolor* se utilizaron

para originar el árbol. Las ramas que conducen a las secuencias de *Ilex* están coloreadas en azul claro, mientras que las ramas que conducen a las secuencias de *Camellia* están coloreadas en verde. Los números sobre las ramas indican el valor de *bootstrap*. **111**

FIGURA 44: Alineamiento de múltiples secuencias de las enzimas tipo CS previamente caracterizadas de *Coffea arabica*, *Coffea canephora*, *Camellia sinensis*, *Citrus sinensis*, *Theobroma cacao* y *Paullinia cupana*, con las tres enzimas CS caracterizadas en este estudio de *Ilex paraguariensis* y un ácido salicílico carboxil metiltransferasa (SAMT) de *Clarkia breweri*. Los gráficos de estructura secundaria están dados para CcDXMT arriba. Los motivos de unión SAM (A, B' y C) y la región conservada YFFF se muestran con cuadros grises. El dominio de unión al sustrato está indicado por los residuos coloreados. Las estrellas amarillas marcan los residuos cruciales para IpCS1 versus IpCS2 y las estrellas rosadas marcan los cruciales residuos para IpCS3 versus IpCS1 e IpCS2. Los números de acceso son los siguientes: TcCS1 (EOY17874), PcCS1 (XXX), IpCS1 (XXX), CisXMT1 (KD050937), TcCS2 (EOY17880), PcCS2 (XXX), IpCS2 (XXX), TCS2 (AB031281), CcXMT1 (DQ422954), CaXMT1 (AB048793), CaXMT2 (JX978522), TCS1 (AB031280), BTS1 (AB096699), CcMXMT1 (JX978517), CaMXMT1 (JX978519), CaMXMT2 (AB084126), CisXMT2 (XP_006469448), PcCS (DAA64605), CaDXMT1 (AB084125), CaDXMT2 (KJ577793), CcDXMT (JX978516), IpCS3 (XXX) y CbSAMT (AAF00108). **112** **115-116**

FIGURA 45: Interacciones clave entre las enzimas tipo CS de *I. paraguariensis* y sus sustratos. Para cada panel, el lado izquierdo muestra la estructura proteína-ligando, mientras que el lado derecho muestra un esquema de interacción simplificado. Los enlaces de hidrógeno se indican como líneas discontinuas entre parejas de interacción en ambos lados. Las interacciones hidrofóbicas se ilustran como líneas de contorno suaves entre los aminoácidos respectivos y el ligando. **118**

- A. IpCS1 - complejo de xantina
- B. IpCS2 - complejo de 3-metilxantina
- C. IpCS3 - complejo de teobromina
- D. IpCS3 - complejo de paraxantina

FIGURA 46: Comparación regiones sintéticas CS. En rectángulos se muestran los genes CS. En líneas negras se denotan las duplicaciones en tándem. **123**

108

Tesis Doctoral
Sebastián Maximiliano Rossi

FIGURA 47: Comparación regiones sintéticas XMT. Comparación de las regiones que contienen al menos uno de los genes participantes de la vía de biosíntesis de cafeína **124**

FIGURA 48: Comparación regiones sintéticas del bloque genómico que contiene los genes XMT de la vía de biosíntesis de cafeína en *Coffea canephora* **125**

FIGURA 49: Comparación regiones genómicas del bloque XMT de *Coffea canephora*. **126**
Se presentan muchas regiones sintéticas de *I. paraguariensis* con *Citrus* y *Coffea*, pero sin genes XMT ortólogos

FIGURA 50: Región sinténica CS: Representación esquemática de los resultados de 127 microsintenia sobre los bloques genómicos CS de la vía metabólica principal de síntesis de cafeína analizados de las especies productoras y no productoras de este alcaloide.

FIGURA 51: Región sinténica XMT: representación esquemática de los resultados de

128

microsintenia sobre los bloques genómicos XMT de la vía principal de síntesis de cafeína analizados de las especies productoras y no productoras de este alcaloide.

FIGURA 52: Se puede observar en este análisis sinténico la ausencia de genes CS en el 129

genoma de *Coffea arabica* demostrando que dicha especie no tuvo cooptación durante su evolución que utilizar solamente la vía XMT para sintetizar cafeína.

FIGURA 53: Representación filogenética de órdenes de angiospermas que han 131 perdido genes CS o XMT o ambos. Las líneas verdes representan los órdenes de plantas que presentan genes de tipo CS y las líneas moradas órdenes que presentan genes de tipo XMT. Las X dependiendo el color representa pérdida de dichos genes.

FIGURA 54: Análisis de la expresión génica de IpCS1-3 según lo indicado por la abundancia relativa de lecturas del transcriptoma YM mapeadas de los transcritos de 134 IpCS1-3. Genes de expresión constitutiva: G3PD, gliceraldehído-3-fosfato deshidrogenasa; EF1A, factor de elongación 1 alfa. Los datos de trabajo obtenidos para dichas condiciones del mismo cultivar Pg CA538/90 junto a las condiciones de estrés del cultivar SI-49 de la provincia de Misiones, Argentina.

Tesis Doctoral
Sebastián Maximiliano Rossi

LISTA DE

TABLAS

PÁGINA

TABLA 3: Porcentajes de cafeína y teobromina extraídos de yerba mate.

TABLA 1: Descripción de los conocimientos generados a partir de datos provenientes del transcriptoma y genoma de la yerba mate y los impactos esperados.

TABLA 2: Porcentajes de recuperación de extracto orgánico conteniendo las metilxantinas y contenido de cafeína y teobromina en porcentaje peso en peso (P/P) determinado por HPLC en hojas de especies de *Ilex*. La especie *I. paraguariensis* var. *paraguariensis* presentó un contenido de cafeína de 0,65% y un 0,12% de teobromina. *Ilex paraguariensis* var. *vestita* presentó un porcentaje menor de cafeína (0,003%) pero un porcentaje mayor de teobromina (estimada en 0,22%). No se detectaron metilxantinas en *Ilex brevicuspis*, *Ilex dumosa* e *Ilex microdonta*. También realizaron estudios en cromatografía en capa fina (TLC), la teofilina no fue detectada en ninguna de las plantas y sólo se halló cafeína y teobromina en las variedades de *Ilex paraguariensis* (Adaptado de Reginatto *et al.* 1999). a. Las soluciones estándar y de muestra se inyectaron por triplicado. b. Teobromina diluida con otra sustancia, la resolución de ambos picos no fue adecuada para la determinación cuantitativa.

TABLA 4: Cuantificación del ARN extraído. Comparación de los valores de absorbancia obtenidos mediante espectrofotometría con la muestra de ARN purificado de *Ilex paraguariensis* cultivar filial Pg538.

TABLA 5: Reactivos y amplificación de RT-PCR

TABLA 6: *Primers* y productos de amplificación

TABLA 7: Reactivos para la amplificación del producto de PCR

TABLA 8: Condiciones de ciclado

TABLA 9: Protocolo de ligación utilizado con incubación de una hora a temperatura ambiente

TABLA 10: Datos de las condiciones de yerba mate analizadas para el cálculo de la abundancia relativa de expresión de las cafeínas sintasas en los transcriptomas disponibles en NCBI.

12-14

70

72

73

74

65

75

77

66

132-133

Tesis Doctoral
Sebastián Maximiliano Rossi
Tesis Doctoral
Sebastián Maximiliano Rossi

RESUMEN

Ilex paraguariensis St. Hil. (familia Aquifoliaceae) es un árbol dioico que es originario de las selvas subtropicales de América del Sur. Tanto sus hojas como sus pequeñas ramas secas son procesadas y utilizadas para preparar la infusión popular, estimulante y de sabor particular, conocida como “mate”, de allí que su nombre coloquial en América Latina sea “yerba mate”. Los estudios sobre las propiedades medicinales de la yerba mate se han incrementado en la última década y existe un consenso actual sobre sus efectos benéficos como antioxidante, antiinflamatorio, antimutagénico y sobre el metabolismo lipídico y de la cafeína que renuevan el interés en el estudio de esta importante especie.

El objetivo general del trabajo se basó en el análisis desde el punto de vista genómico y filogenético de la vía biosintética de la cafeína comparativamente con la existente en otros organismos de importancia agronómica como *Coffea*, *Theobroma*, *Camellia*, entre otras, prediciendo el impacto de las variaciones encontradas a nivel proteómico y metabolómico. Con el propósito de elucidar el repertorio de expresión de los genes relacionados a la vía biosintética de la cafeína de yerba mate, se exploró y descubrió una vasta colección de transcriptos de *I. paraguariensis* utilizando como genoma de referencia a la línea Pg538, material elite desarrollado por el INTA EAA-Cerro Azul de Misiones y utilizando los datos aportados por el transcriptoma de *I. paraguariensis* secuenciado por *Debat et. al* en 2014. De

este análisis, a partir las 72.031.388 secuencias de 100 pb obtenidas y las lecturas de alta calidad, fueron ensambladas *de novo* en 44.907 transcritos, abarcando 40 millones de bases con una cobertura estimada de 180X. Múltiples análisis de secuencias permitieron inferir que la yerba mate contiene alrededor de 32.355 genes y 12.551 variantes. Asimismo, mediante la herramienta del transcriptoma, se identificaron y categorizaron transcritos pertenecientes a más de 100 vías metabólicas, y gracias a esta primera descripción, se procedió a realizar el análisis genómico comparativo y filogenético de la vía metabólica de los transcritos involucrados en producción de la cafeína.

Se realizó un examen filogenético de las enzimas de una gran familia de genes denominados SABATH (Salicylic Acid, Benzoic Acid, Theobromine methyltransferase) que catalizan la metilación de átomos de oxígeno de una amplia diversidad de ácidos carboxílicos como, por ejemplo, ácido antranílico, benzoico, giberílico, jasmónico, logánico, salicílico y ácido indol-3-



Universidad Nacional de Misiones
Facultad de Ciencias Exactas, Químicas y Naturales



Página | 1

Tesis Doctoral
Sebastián Maximiliano Rossi

acético. Los genes evolucionados más recientemente de la familia SABATH son los que corresponden a la vía XMT (Metiltransferasas de Xantina) y CS (Cafeína Sintasa). Se obtuvieron 19 enzimas SABATH del transcriptoma y, a partir de los datos generados del genoma se analizaron 26 genes completos y 3 pseudogenes, para determinar qué tipo de metiltransferasas utiliza la yerba mate en el proceso metabólico de la conversión de cafeína.

La vía de biosíntesis de la cafeína es una vía común a una serie de especies de gran importancia económica por la generación de teobromina y cafeína como bioproductos de gran impacto económico y cultural. Se pudieron identificar, categorizar y caracterizar tres genes que codifican para las enzimas que catalizan los 3 pasos de reacción de la vía de biosíntesis desde la xantina a la conversión de cafeína en yerba (IpCS1, IpCS2 e IpCS3). Mediante ensayos enzimáticos de los productos purificados, pudieron detectarse los sustratos específicos midiendo actividad catalítica de cada enzima. Se los nombraron como IpCS1, IpCS2 e IpCS3 ya que siguen la vía de conversión mediada por cafeína sintasa (CS).

Se concluyó que la enzima (IpCS1) muestra preferencia para catalizar la formación de 3-metilxantina a partir de xantina como sustrato. La segunda enzima descrita en este estudio (IpCS2) presenta actividad con 3-metilxantina para producir teobromina, pero no presenta actividad alguna con otros sustratos. La tercera enzima, (IpCS3) metila al máximo teobromina para formar cafeína. Colectivamente estas tres enzimas son capaces de catalizar una vía

completa de biosíntesis desde xantina a cafeína en yerba mate. Dicha vía desde X>>3X>>TB>>CF ha sido también descrita para *Theobroma* y *Paullinia* y éstas plantas también usan enzimas SABATH de tipo CS para catalizar las reacciones de producción de cafeína.

El conocimiento de la genómica comparativa de las enzimas involucradas en esta vía permitió cuantificar el peso de las variaciones genéticas de la yerba mate en comparación con otras especies productoras de cafeína de gran interés agronómico mediante un estudio de microsintenia y al mismo momento se compararon órdenes de plantas no productoras de cafeína, pudiendo denotarse la falta de éstos genes ortólogos de una vía que ha sido muy conservada en la evolución de varias especies de eudicotas.

PALABRAS CLAVES: Biosíntesis de Cafeína; Genómica; Cafeína sintasa; *Ilex paraguariensis*; SABATH; Metiltransferasas de Xantina



Universidad Nacional de Misiones
Facultad de Ciencias Exactas, Químicas y Naturales



Página | 2

Tesis Doctoral
Sebastián Maximiliano Rossi

ABSTRACT

Ilex paraguariensis St. Hil. (Aquifoliaceae family) is a dioecious tree native from the subtropical forests of South America. Its leaves and small twigs processed are used to prepare the popular, stimulating and particular flavor infusion, known as "mate", hence its colloquial name in Latin America be "yerba mate". Studies on the medicinal properties of yerba mate have increased in the last decade and there is a current consensus on its beneficial effects as an antioxidant, anti-inflammatory, anti-mutagenic and on lipid and caffeine metabolism that renews the interest of the study of this important specie.

The general objective of the work was based on the analysis from the genomic and phylogenetic point of view of the caffeine biosynthetic pathway in comparison with other organisms of agronomic importance, such as *Coffea*, *Theobroma*, *Camellia* among others, predicting the impact of changes encountered at proteomic and metabolomics level. With the aim of elucidating the repertoire of yerba mate gene expression related to the caffeine biosynthetic pathway, it was explored and discovered a vast collection of transcripts of *I. paraguariensis* using as a reference genome the Pg538 line, elite material developed by INTA EAA-Cerro Azul Misiones.

This could be carried out through the data provided by the transcriptome of *I. paraguariensis* sequenced by *Debat et. al* in 2014, obtaining 72,031,388 sequences of 100 bp. The high quality readings obtained were assembled *de novo* in 44,907 transcripts, covering 40 million bases with an estimated 180X coverage. Multiple sequence analysis allowed us to infer that the yerba mate contains ~ 32,355 genes and 12,551 gene variants.

Also, they were identified and categorized transcripts from more than 100 metabolic pathways and thanks to this first description, we proceeded to perform the comparative and phylogenetic genomic analysis of the metabolic pathway of the transcripts involved in the caffeine production.

A phylogenetic examination of the enzymes of a large family of genes named SABATH (Salicylic Acid, Benzoic Acid, Theobromine methyltransferase) that catalyze the methylation of oxygen atoms of a wide variety of carboxylic acids such as, for example, anthranilic, benzoic, giberilic acid, jasmonic, loganic, salicylic and indole-3-Acetic Acid, was performed. The most recent



Universidad Nacional de Misiones
Facultad de Ciencias Exactas, Químicas y Naturales



Página | 3

Tesis Doctoral
Sebastián Maximiliano Rossi

evolved SABATH genes correspond to the XMT (Xanthosine Methyltransferase) and CS (Caffeine Sintase) enzymes pathway. 19 transcriptome SABATH enzymes were obtained, and 26 complete genes and 3 pseudogenes were analyzed from the genome generated data, to determine what type of methyltransferases are used in the metabolic process of caffeine carried out by yerba mate.

The caffeine biosynthetic pathway is a common pathway to a number of species of great economic importance for the generation of theobromine and caffeine as bioproducts of great economic and cultural impact. The three genes encoding the enzymes that catalyze the 3 reaction steps of the biosynthesis pathway from xanthine to caffeine conversion could be identified (IpCS1, IpCS2 and IpCS3). By enzymatic assays of the purified products, specific substrates could be detected by measuring catalytic activity of each enzyme. They were named as IpCS1, IpCS2 and IpCS3 as they follow the caffeine synthase (CS) mediated conversion pathway.

It concluded, one enzyme (IpCS1) appears to prefer to catalyze the formation of 3-methylxanthine when provided xanthine. A second enzyme (IpCS2), shows activity with 3-methylxanthine to produce theobromine but no activity with any other substrate. A third

1.1.5. BIOSÍNTESIS DE CAFEÍNA EN *Citrus sinensis*

(NARANJA) 21

1.1.6. CAFEÍNA SINTASA DE *Ilex paraguayensis* 23

1.2. HIPÓTESIS Y OBJETIVOS 25 1.2.1. OBJETIVOS ESPECÍFICOS 25 1.3.
HITOS DE EVALUACIÓN 26

CAPÍTULO 2

2. MARCO TEÓRICO 27 2.1. GENÉTICA DE LA YERBA MATE 28 2.2.
IMPORTANCIA AGRONÓMICA 30 2.3. IMPORTANCIA DE LA VÍA DE
BIOSÍNTESIS DE CAFEÍNA 33 2.4. PRODUCCIÓN DE CAFEÍNA EN
ANGIOSPERMAS 33 2.5. GENES CONSERVADOS EN LA EVOLUCIÓN 34

Página | 5

Tesis Doctoral
Sebastián Maximiliano Rossi

2.6. EVOLUCIÓN DE VÍAS METABÓLICAS 35 2.6.1. RESURRECCIÓN DE
SECUENCIAS ENZIMÁTICAS
ANCESTRALES 39 2.6.2. EVOLUCIÓN CONVERGENTE DE
CAFEÍNAS
SINTASAS EN RÓSIDAS Y ASTÉRIDAS 42 2.6.3.
MECANISMOS DE DUPLICACIÓN Y FUSIÓN
GÉNICAS 43 2.7. VÍA METABÓLICA DE LA CAFEÍNA 47 2.7.1. LAS
XANTINAS 47 2.7.2. METILXANTINAS 48 2.7.2.1. DISTRIBUCIÓN 48 2.7.3.
ENZIMAS DE LA FAMILIA SABATH (SALICYLIC
ACID, BENZOIC ACID, THEOBROMINE
METHYLTRANSFERASE) 48 2.7.3.1. CARACTERÍSTICAS Y
DISTRIBUCIÓN 50 2.8. GENERACIÓN DE LA VÍA DE BIOSÍNTESIS DE
CAFEÍNA EN PLANTAS SUPERIORES 52 2.8.1. FORMACIÓN DE 7-
METILXANTINA 52 2.8.2. FORMACIÓN DE TEOBROMINA 53 2.8.3.
CONVERSIÓN DE TEOBROMINA EN CAFEÍNA 54 2.8.4. SUMINISTRO DE
XANTOSINA PARA LA
BIOSÍNTESIS DE CAFEÍNA 55 2.8.4.1. RUTA *DE NOVO* 55
2.8.4.2. RUTA AMP 56 2.8.4.3. RUTA SAM 56 2.8.4.4. RUTA GMP
57
2.8.5. N-METILTRANSFERASAS INVOLUCRADAS EN LA BIOSÍNTESIS
DE METILXANTINA 57 2.8.5.1. EXPRESIÓN GÉNICA EN PLANTAS DE CAFÉ
Y TÉ 57 2.8.5.2. RELACIÓN EVOLUTIVA DE LA CAFEÍNA SINTASA Y

ENZIMAS RELACIONADAS 58 2.9. LA CAFEÍNA 60 2.9.1. PROPIEDADES QUÍMICAS 60 2.9.2. MECANISMOS DE ACCIÓN 62 2.9.3. CAFEÍNA EN EXTRACTOS ACUOSOS 64 2.9.3.1. IMPORTANCIA PRODUCTIVA DEL CONOCIMIENTO DE LOS VALORES DE METILXANTINAS EN YERBA MATE 64

Página | 6

Tesis Doctoral
Sebastián Maximiliano Rossi

2.10. ANÁLISIS EVOLUTIVOS 66 2.10.1. SINTENIA Y MICROSINTENIA 66

CAPÍTULO 3

3. MATERIALES Y MÉTODOS 68 3.1. MATERIAL DE ESTUDIO 68 3.2. EXTRACCIÓN DE ÁCIDOS NUCLEICOS
3.2.1. EXTRACCIÓN DE ADN GENÓMICO 69 3.2.2. TÉCNICAS EXPERIMENTALES CON ARN 69 3.2.2.1. EXTRACCIÓN DE ARN CELULAR TOTAL 70 3.3. CUANTIFICACIÓN ESPECTROFOTOMÉTRICA DEL ARN 70 3.3.1. CÁLCULO DE LOS VALORES DE CONCENTRACIÓN E ÍNDICES DE ABSORBANCIA 71 3.4. CÁLCULO DEL VOLUMEN FINAL UTILIZADO 71 3.5. REACCIÓN DE RETROTRANSCRIPCIÓN PCR (RT-PCR) 71 3.6. AMPLIFICACIÓN DE ADNC MEDIANTE LA REACCIÓN EN CADENA DE LA POLIMERASA 72 3.6.1. AMPLIFICACIÓN DE LA REGIÓN CORRESPONDIENTE A IpCS1 72 3.7. ELECTROFORESIS EN GELES DE AGAROSA 75 3.7.1. PROCEDIMIENTOS 76 3.8. EXTRACCIÓN EN GEL DE AGAROSA 76 3.9. LIGACIÓN Y TRANSFORMACIÓN 76 3.10. EXTRACCIÓN DE PLÁSMIDOS 78 3.10.1. INDUCCIÓN DE EXPRESIÓN 80 3.11. SECUENCIACIÓN DE LOS FRAGMENTOS DE INTERÉS OBTENIDOS 80 3.12. EXPRESIÓN HETERÓLOGA Y PURIFICACIÓN DE ENZIMAS 80 3.12.1. EXPRESIÓN DE PROTEÍNAS 80 3.12.2. PURIFICACIÓN DE PROTEÍNAS 81 3.13. ANÁLISIS FUNCIONAL IN VITRO 81 3.14. ESTIMACIÓN DE LOS PARÁMETROS DE MICHAELIS-MENTEN 82 3.15. CROMATOGRAFÍA LÍQUIDA DE ALTO RENDIMIENTO 82 3.16. CROMATOGRAFÍA LÍQUIDA - ESPECTROMETRÍA DE MASAS EN TÁNDEM 83

3.17. BIOINFORMÁTICA	84	3.18. ANÁLISIS FILOGENÉTICO	84	3.19. ENSAYO ENZIMÁTICO – ESQUEMA GENERAL	85
3.20. GENÓMICA COMPARATIVA	85	3.20.1. ANÁLISIS DE MICROSINTENIA	85	3.21. MODELADO MOLECULAR	86

CAPÍTULO 4

4. RESULTADOS Y DISCUSIÓN	88	4.1. ANÁLISIS DE LOS GENES INVOLUCRADOS EN LA VÍA DE BIOSÍNTESIS	88
4.2. SECUENCIACIÓN DEL GEN CS CANDIDATO	89	4.3. ENSAYO ENZIMÁTICO DE LAS PRIMERAS SECUENCIAS IDENTIFICADAS	90
4.4. SÍNTESIS Y CLONACIÓN DE SECUENCIAS #2 Y #4 DEL GENOMA	97	4.5. PSEUDOGENES CARACTERIZADOS EN LA FAMILIA SABATH DE YM	98
4.6. RECONSTRUCCIÓN DE LA VÍA COMPLETA DE BIOSÍNTESIS DE CAFEÍNA EN <i>I. paraguariensis</i>	99	4.6.1. ENSAYO ENZIMÁTICO DE LAS DOS ENZIMAS CS RESTANTES	99
4.6.2. ANÁLISIS DESCRIPTIVO DE LA BIOSÍNTESIS DE CAFEÍNA EN YM	102	4.6.3. LOCALIZACIÓN DE LOS GENES DE LA VÍA BIOSINTÉTICA DE LA CAFEÍNA EN YM	105
4.6.4. ANÁLISIS COMPARATIVO DE <i>Camellia</i> , <i>Theobroma</i> , <i>Paullinia</i> Y <i>Coffea</i>	106	4.7. CONSTRUCCIÓN DEL ÁRBOL FILOGENÉTICO DE ENZIMAS SABATH DE YM CON OTRAS ESPECIES DE INTERÉS PRODUCTIVO	108
4.8. CARACTERIZACIÓN BIOINFORMÁTICA DE LOS GENES DE LA VÍA BIOSINTÉTICA	113	4.8.1. EL MODELADO COMPUTACIONAL DE SITIOS ACTIVOS IpCS MUESTRA SUSTITUCIONES	

DE AMINOÁCIDOS CONVERGENTES 114 4.9. ANÁLISIS
EVOLUTIVOS 119 4.9.1. MICROSINTENIA 119 4.9.2. LA GENÓMICA
COMPARATIVA EVOLUTIVA

DE LA BIOSÍNTESIS DE CAFEÍNA EN LAS
ANGIOSPERMAS REVELA RESTRICCIONES PARA
LA COOPCIÓN DE GENES CONVERGENTES 130 4.10.
ABUNDANCIA RELATIVA DE TRANSCRITOS lpcSs DE

YERBA MATE 131

CAPÍTULO 5

**5. CONCLUSIONES 135 6. PERSPECTIVAS FUTURAS 138 7. MATERIAL
SUPLEMENTARIO 139 8. REFERENCIAS BIBLIOGRÁFICAS 194**

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. ANTECEDENTES

Teniendo en cuenta los estudios de genética poblacional de yerba mate, la heterogeneidad genética encontrada en plantaciones de origen brasileño (Neumann-Wendt 2005) es tan elevada como la variación poblacional natural de la especie (Gauer & Cavalli-Molina 2000), probable reflejo de la superposición casi total entre la zona de producción de los tres países yerbateros (Argentina, Brasil y Paraguay) y el área de distribución natural. Por su parte Belingheri & Prat Kricun (1997) reportan que, en yerbales en plena producción, menos del 35% de las plantas aportan más del 50% de la producción. Estos inconvenientes han suscitado la implementación de diversas estrategias de mejoramiento genético de la yerba mate tendientes a reducir la variabilidad no deseada, por ejemplo se puede destacar, la selección de clones y progenies de alto rendimiento, calidad y adaptabilidad ambiental (Belingheri & Prat Kricun 1997), la regeneración de plantas mediante cultivo *in vitro* (Sansberro *et al.* 2004; Rey *et al.* 2002; Tarragó *et al.* 2005), la obtención y distribución a principio de la década del 90 de semillas policlonales y biclonales, los ensayos de propagación vegetativa (INTA 2010; Bubillo *et al.* 2011, 2012) entre otros. Sin embargo, aparte de este carácter, probablemente poligénico o aditivo, se sabe muy poco acerca de otros *loci* de importancia agronómica.

El mejoramiento genético de la yerba mate es clave para el ascenso de las economías locales de la provincia de Misiones, íntimamente relacionadas con el cultivo, las cuales se se verían beneficiadas con el valor agregado que supondría la introducción de cultivares de origen certificado y *blend*, como por ejemplo, con mejores características organolépticas relacionadas con el sabor (asociado con una variación en los contenidos de cafeína, teobromina, glucosa y sacarosa) y con el aroma (asociado con la presencia de más de 40 compuestos volátiles). Asimismo, los productores serían beneficiados con la introducción de plantas con una mayor uniformidad en la densidad de ramas, tiempo de brotación, rendimiento unitario y tolerancia a plagas específicas de la yerba mate, como el psílido, taladro grande, marandová, como también con variedades adaptadas a diferentes regiones y

exigencias, como la tolerancia a períodos de estrés hídrico, bajas temperaturas, heladas y alta insolación. Dada la importancia multifactorial del cultivo de la YM (Yerba Mate) se llevó a cabo una investigación del transcriptoma publicado en el año 2014 mostrando una primera aproximación de los genes involucrado en varias vías metabólicas como la predicción de sus enzimas y propiedades (Debat *et al.* 2014).

En el trabajo de investigación de Debat *et al.*, la obtención de una progenie segregante de mapeo y la construcción de un mapa genético preliminar de *I. paraguariensis* facilitó la detección de *loci* asociados a caracteres de interés agronómico como así también la rápida introgresión de determinadas regiones génicas a los materiales élite (hojas de yerba mate). Dos genotipos genéticamente divergentes con comportamiento contrastante respecto a la tolerancia a sequía se seleccionaron como progenitores femenino y masculino. Así también mediante cruzamientos controlados y rescate de embriones, los autores establecieron una población F₁ tipo “pseudo-test cross” de 700 individuos, con potencial para extenderse hasta 1900, donde segregan tanto la tolerancia a sequía como el género. Un subgrupo de 117 plantas fue empleado para construir un mapa genético preliminar. En dicho proyecto se ensayaron 5 combinaciones de cebadores de AFLP y 16 cebadores de RAPDs, para generar 119 marcadores informativos, de los cuales el 68,9% se ajustó a los valores de segregación mendeliana esperados. Los marcadores originados en cada progenitor se analizaron independientemente para generar un mapa femenino formado por 11 grupos de ligamiento y otro masculino con 16 grupos. El sexado de las plantas de la F₁, su clasificación en rangos de tolerancia y la densificación del mapa mediante el mapeo de marcadores adicionales facilitó la detección de marcadores asociados a caracteres de gran interés para el cultivo de yerba (Debat *et al.* 2014).

Con el objetivo de elucidar el repertorio de expresión de los genes de yerba mate, los autores exploraron y descubrieron una vasta colección de transcriptos de *I. paraguariensis* utilizando como genoma de referencia a la línea Pg538, material elite desarrollado por el INTA EAA-Cerro Azul de Misiones, mediante el uso de la tecnología de secuenciación paralela masiva de ARN (RNA-seq) de nueva generación (*Next Generation Sequencing*, NGS). El ARN total de *I. paraguariensis* fue secuenciado con la plataforma Illumina HiSeq™-2000 obteniéndose 72.031.388 secuencias de 100 pb. Las lecturas obtenidas de alta calidad fueron ensambladas *de novo* en 44.907 transcriptos, abarcando 40 millones de bases con una cobertura estimada

de 180X. Múltiples análisis de secuencias permitieron inferir que la yerba mate contiene alrededor de 32.355 genes y 12.551 variantes de genes. En líneas generales, se identificaron y categorizaron transcritos pertenecientes a más de 100 vías metabólicas. Asimismo, se identificaron en el análisis del transcriptoma alrededor de 1.000 factores de transcripción putativos, genes implicados en estrés por calor y estrés oxidativo, respuesta a patógenos, resistencia a enfermedades y respuesta a hormonas. También se identificaron transcritos relacionados a estrés osmótico, sequía, salinidad, estrés por frío, senescencia, floración temprana y determinación sexual (Debat *et al.* 2014). Este trabajo constituye un antecedente en la caracterización de la actividad de un gen central del metabolismo en respuesta a una situación de estrés de ocurrencia frecuente en los campos de cultivo de yerba mate.

En la tabla 1 se resumen los diversos impactos que se potencian a partir de los datos generados del transcriptoma y el genoma, lo que a su vez permite expandir el uso de los productos agrícolas hacia fines biomédicos, químicos y nutricionales.

TABLA 1: Descripción de los conocimientos generados a partir de datos provenientes del transcriptoma y genoma de la yerba mate y los impactos esperados. (Datos tomados del proyecto Promatear)

CONOCIMIENTO GENERADO	IMPACTO
Genes y variantes génicas implicadas en la metabolización de compuestos nitrogenados y la interacción micorriza-planta.	Receptividad a los programas de fertilización y el uso de consorcios de microorganismos como biofertilizantes.

<p>Genes implicados en las vías metabólicas secundarias, dentro de éstos se pueden destacar:</p> <p>Genes y variantes génicas implicadas en la generación de compuestos polifenólicos (flavonoides y derivados cafeilquínicos).</p>	<p>Producción de fitoquímicos de interés agroquímico, alimenticio, farmacéutico, cosmético.</p> <p>Potencialidad para la selección de variantes elite para la producción de productos farmacéuticos antioxidantes con diversas acciones farmacológicas para el tratamiento de enfermedades cardiovasculares, tumorales neuro degenerativas y demás enfermedades asociadas al estrés</p>
---	---

<p>Genes y variantes génicas de las vías productoras de alcaloides.</p> <p>Genes y variantes génicas de las vías productoras de saponinas.</p>	<p>oxidativo. Productos anorexígenos. Cosméticos antiage, protectores solares. En nutracéutica.</p> <p>Potencialidad para la selección de variantes con alto contenido en alcaloides élite para la producción de bebidas energizantes, yerba fuerte, productos alimenticios combinados, productos farmacéuticos estimulantes del sistema nervioso central y respiratorio, diuréticos, antimigrañosos, contra los mareos, la obesidad.</p> <p>Productos cosméticos anticelulíticos, lipolíticos.</p> <p>Productos con bajo contenido de alcaloides, yerba suave. Potencialidad para la selección de variantes para la producción de productos farmacéuticos hipocolesterolemiantes.</p>
--	--

Genes y variantes génicas implicados en la resistencia a factores ambientales.	Potencialidad biotecnológica en la generación de insumos para ingeniería genética (obtención de promotores y genes de respuesta a factores ambientales destinados a la producción de vectores de expresión y/o el uso en la producción de recombinantes).
Genes y variantes génicas implicados en la resistencia a insectos.	Potencialidad para seleccionar individuos resistentes. Potencialidades biotecnológicas en la generación de insumos para ingeniería genética.
Genes y variantes génicas implicados en la resistencia a hongos fitopatógenos.	Potencialidad para seleccionar individuos resistentes. Potencialidades biotecnológicas en la generación de insumos para ingeniería genética.
Genes y variantes génicas implicadas en el desarrollo temprano y el enraizamiento.	Potencialidad para seleccionar material elite para un mayor rendimiento en vivero.

Genes implicados en la determinación sexual.	Potencialidad para seleccionar género a nivel de plántula.
Genes y variantes génicas relacionadas al desarrollo estacional.	Potencialidad para seleccionar material elite con mayor producción de hoja por Ha.
Genes y variantes génicas relacionadas a la producción de clorofila y otros pigmentos.	Producción de pigmentos. Respuesta a la luz.

transcriptómica de *I. paraguariensis* permitió comprender con profundidad su bioquímica y asimismo evaluar el impacto de las variantes alélicas que se presentan sobre la metabolómica de la especie generando la capacidad de poder implementar mejores técnicas de selección de individuos plus que muestran un balance metabólico determinado, como por ejemplo una mayor o menor producción de cafeína, basándose en las características proteómicas de las enzimas que forman parte de la vía. Esta posibilidad podría impactar no solo en el conocimiento pormenorizado de las vías de producción de fitoquímicos y productos con interés farmacológico, que permita compararlas con las existentes en otras plantas modelo, sino cambiar la base de selección de individuos destacados basándose en la búsqueda de variantes alélicas basadas en modificaciones sustanciales en puntos clave de las enzimas que otorgan a cada una propiedad fisicoquímica deseada.

1.1.1. PRESENCIA DE CAFEÍNA EN OTRAS ESPECIES DE INTERÉS AGRONÓMICO

En el año 2000 se daba a conocer el gen que codifica la cafeína sintasa (CS) originalmente clonados de hojas jóvenes de té dados por las investigaciones realizadas por Kato *et al.* (2000). Este grupo de investigadores clonaron el gen que codifica la cafeína sintasa (TCS) de las hojas jóvenes de las plantas de té utilizando la técnica de amplificación rápida de extremos de ADN complementarios.

Este gen identificado tiene un tamaño de 1438 pb y codifica para 369 aminoácidos.

Utilizando como secuencia modelo la CS de *Camellia sinensis* se identificó una secuencia de longitud completa de ARNm correspondiente al transcrito del transcriptoma de la yerba mate.

La proteína predictiva de la YM de 366 aminoácidos presentó un 61 % de similitud al correspondiente de *Camellia* de 369 aminoácidos que corresponde al dominio de la familia “Methyltransferase_7 SAM dependent carboxyl methyltransferase”. La predicción de la estructura de la proteína CS corresponde a la estructura cuaternaria de la CS de YM que fue vislumbrada con el servidor SwissModel utilizando como templado la estructura completa de *Coffea arabica* (Debat *et. al.* 2014).

Para poder descubrir los genes y las vías utilizadas por las plantas modernas para sintetizar cafeína, se reportaron análisis bioinformáticos y filogenéticos para revelar que tanto el cacao (*Theobroma cacao* - Tc - Malvales) como la guaraná (*Paullinia cupana* - Pc - Sapindales) expresan múltiples secuencias de tipo CS en sus hojas y/o frutos que son ortólogas a las utilizadas por el té (*Camellia sinensis* - Ericales) en hojas y brotes. Sin embargo, contrario a las expectativas de *Camellia*, la expresión heteróloga y los ensayos de las enzimas CS de *Theobroma* y *Paullinia* indican que catalizan una ruta diferente para sintetizar alcaloides de xantina (Angelo P.C. *et al.* 2008; Argout X. *et al.* 2008). (Fig.1).

Específicamente, ambas especies poseen una enzima (CS1) que preferentemente metila la xantina para producir 3-metilxantina, así como una segunda enzima (CS2) que preferentemente metila la 3-metilxantina para producir teobromina. De manera sorprendente, a pesar de que las enzimas son parte del linaje CS, TcCS1 (*Theobroma cacao* CS1) está más estrechamente relacionado con TcCS2 (*Theobroma cacao* CS2) en lugar de la enzimáticamente similar PcCS1 (*Paullinia cupana* CS1) que, a su vez, está más estrechamente relacionado con PcCS2 (*Paullinia cupana* CS2) (Huang *et al.*, 2016).

FIG.1: La cafeína ha evolucionado de manera convergente en cinco especies de plantas con flores utilizando diferentes combinaciones de genes y vías. (A) Las relaciones filogenéticas entre órdenes de Rósidas y Astéridas muestran múltiples orígenes de la biosíntesis de cafeína. Los linajes de color verde lima rastrean el antiguo linaje CS de enzimas que ha sido reclutadas de forma independiente para su uso en tejidos que acumulan cafeína en *Theobroma*, *Paullinia* y *Camellia*. Linajes turquesa trazan el antiguo linaje XMT que fue reclutado independientemente en *Citrus* y *Coffea*. (B y C) *Theobroma* y *Paullinia* han convergido en vías biosintéticas similares catalizadas por enzimas de tipo CS. (D) *Citrus* ha evolucionado una vía diferente catalizada por enzimas de tipo XMT, a pesar de su estrecha relación con *Paullinia*. (E y F) *Camelia* y *Coffea* catalizan la misma vía utilizando diferentes enzimas. Las vías bioquímicas propuestas se basan en actividades enzimáticas relativas mostradas por gráficos de barras correspondientes que indican actividades relativas medias (de 0 a 1) con ocho sustratos de alcaloides de xantina. CisXMT1 y TCS1 catalizan más de una reacción en las vías propuestas. XMT y CS se han duplicado recientemente e independientemente en cada uno de los cinco linajes. **Fuente:** Los datos fueron tomados con permiso del autor del trabajo publicado “*Convergent evolution of caffeine in plants by co option of exapted ancestral enzymes*, Huang et al, 2016, colaborador directo del presente trabajo doctoral.

cafeína ha evolucionado al menos 5 veces en diferentes órdenes durante la historia de las plantas con flores como se describió previamente. El laboratorio Barkman – Michigan, EEUU, colaboradores directos con el presente trabajo de investigación doctoral, realizó investigaciones con técnicas de resucitación de enzimas ancestrales. Se resucitaron y caracterizaron funcionalmente enzimas ancestrales tipo CS y XMT en cada especie productora de cafeína (*Camellia*, *Coffea*, *Citrus* y *Paullinia*), así como enzimas antiguas CS y XMT compartidas por rósidas y astéridas para vislumbrar en qué momento surgió la capacidad de sintetizar cafeína. Luego, mutaron secuencias CS ancestral en *Paullinia* y XMT ancestral en *Citrus* hacia sus descendientes modernos, respectivamente, para dilucidar cómo se ensamblaba la ruta biosintética de cafeína en cada especie. La primera hipótesis es que la cafeína fue producida por plantas antiguas, en las que todos los CS y XMT ancestrales podían metilar todos los precursores de cafeína en una vía biosintética de cafeína, y estas funciones ancestrales fueron heredadas y divididas en sus descendientes después de duplicaciones de genes en cada especie productora de cafeína. La segunda hipótesis es que la cafeína no fue producida por plantas antiguas, en las que todos los CS y XMT ancestrales se utilizaron para otras funciones además de metilar alcaloides de xantina, y la capacidad de producir cafeína evolucionó independientemente en una pequeña porción de linajes de plantas recientemente.

1.1.2. BIOSÍNTESIS DE CAFEÍNA EN *Coffea canephora* y *C. arabica* (Café)

La cafeína se sintetiza en las hojas de café, donde tiene propiedades insecticidas (Nathanson, 1984), y en las frutas y semillas, donde también inhibe la germinación de las especies competidoras (Pacheco *et al.* 2008). Los últimos pasos en la biosíntesis de cafeína están mediados por una serie de metiltransferasas (MT) de la familia SABATH de tipo XMT.

Una sola familia de genes (ORTHOMCL170) agrupa 23 genes en el café, pero ninguno en la vid, tomate o *Arabidopsis* y este grupo contiene genes que codifican enzimas conocidas de la ruta biosintética de cafeína (McCarthy *et al.* 2007; Ogawa *et al.* 2001). El análisis filogenético de máxima similitud de ORTHOMCL170 con MT de té y cacao que tienen actividades similares revela clados genéticos específicos de la especie.

Se analizaron esas relaciones de especies en un contexto evolutivo más amplio al incluir muestras de metiltransferasas de todo el genoma de café, cacao y otras especies de eudicotas. Los árboles de máxima similitud muestran que los genes que codifican los parientes MT de *Arabidopsis* más cercanos de las enzimas biosintéticas de cafeína del café están involucrados en las funciones benzoicas, salicílicas y nicotínicas. Las MT biosintéticas de cafeína del café anidados dentro de un clado génico distinto de los del cacao o el té, se agrupan como linajes hermanos.

Respecto a la vía biosintética de la cafeína en el café tres pasos de metilación son necesarios para producirla desde la xantosina, que involucra la acción sucesiva de tres MT: xantosina metiltransferasa (XMT), teobromina sintasa [7-metilxantina metiltransferasa (MXMT)] y cafeína sintasa [3,7-dimetilxantina metiltransferasa (DXMT)].

Los análisis de microsintenia del bloque ORTHOMCL170, que incluye tres sets en tándem, muestran que algunos genes de cafeína sintasa conocidos como: CcXMT, CcMTL y CcNMT3 forman un conjunto de duplicados en tándem coexpresados que forman parte de un clúster de genes metabólicos reminiscentes (Field y Osbourn, 2008; Matsuno *et al.* 2009).

Dado que algunos clústeres de genes metabólicos de plantas son de origen relativamente reciente, muchos estudios se centraron en buscar desentrañar aún más el papel de la duplicación de genes en la expansión de la familia de genes MT del café. Los tres clados principales de café MT en ORTHOMCL170 se distribuyen entre un mínimo de tres bloques genómicos, sin embargo, algunos duplicados en tándem recientes filogenéticamente, se han alejado de sus posiciones originales a través de reordenamientos de los bloques genómicos. Uno de esos movimientos que involucra el supuesto clúster metabólico parece haber dejado atrás el gen CcDXMT (que codifica la 3,7-dimetilxantina metiltransferasa), físicamente separado de su set en tándem ancestral.

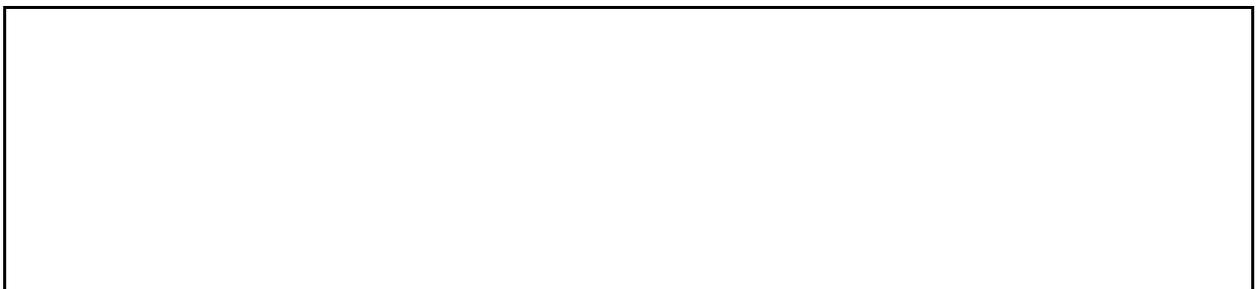


FIG.2: Biosíntesis de cafeína en *Coffea arabica*. Ésta vía de biosíntesis se lleva a cabo por enzima de la familia SABATH de tipo XMT. El primer paso de formación de 7-metilxantina está dado por CaXMT a partir de Xantosina. Para formar teobromina, otra enzima la CaMXMT cataliza dicha formación a partir de 7-metilxantina y CaDXMT es la responsable del último paso de la vía para formar cafeína a partir de teobromina. **Fuente:** Adaptado de Huang, R. "Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants" (2017).

1.1.3. BIOSÍNTESIS DE CAFEÍNA EN *Camellia sinensis* (TÉ)

Como pariente cercano de la enzima biosintética de cafeína TCS1, la enzima TCS2 se clonó a partir de hojas de *Camellia*. El análisis filogenético de ambos genes con otros miembros de SABATH indica claramente que TCS2 es un candidato razonable para el primer paso de metilación de la vía biosintética de cafeína en ésta especie, porque está estrechamente relacionado con todas las demás enzimas CS (Yoneyama *et al.* 2006). Además, la cafeína se encuentra principalmente en las hojas y brotes de *Camellia sinensis* y tanto TCS1 como TCS2 tienen muchas secuencias EST (Expressed Sequence Tag) representadas exclusivamente en las hojas, mientras que otras secuencias SABATH, no. La enzima TCS2 expresada de forma heteróloga tiene la mayor actividad relativa con la xantosina, lo que sugiere que es una 7-metilxantosina sintasa. El producto metilado 7-metilxantosina aún no ha sido verificado por HPLC porque no hay 7-metilxantosina disponible en el mercado para usar como estándar. Sin embargo, un pico de 7-metilxantina aparece en la traza de HPLC de TCS2 en extracto crudo incubado con una mezcla de reacción de xantosina y falta en el ensayo de control negativo sin la adición de xantosina. Es posible que la 7-metilxantosina no sea un compuesto estable, por lo cual se descompone en 7-metilxantina cuando la mezcla de reacción se acidifica para el análisis por HPLC o mediante la acción de otras enzimas bacterianas en el extracto crudo que separa el grupo ribosa de 7- metilxantosina. Una tercera posibilidad es que TCS2 puede ser una enzima de doble función que también es responsable de escindir el grupo ribosa de la 7-metilxantosina, porque el mismo resultado también se encuentra en el análisis por HPLC del producto de reacción de *Coffea canephora* XMT purificado con xantosina (McCarthy *et al.* 2007). Sin embargo, se puede concluir que TCS2 es capaz de metilarse en la posición N-7 de la xantosina como se esperaba y es la cafeína sintasa que faltaba para catalizar el primer paso en la vía biosintética de la cafeína en *Camellia* (fig.3).

FIG.3: Biosíntesis de cafeína en *Camellia sinensis*. La enzima TCS2 cataliza el primer paso de la vía convirtiendo xantosina a 7-metilxantina, mientras que la enzima TCS1 tiene una acción dual sobre 7-metilxantina y teobromina para formar cafeína. **Fuente:** Adaptado de Huang, R. "Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants" (2017).

1.1.4. BIOSÍNTESIS DE CAFEÍNA EN *Paullinia cupana* (GUARANÁ)

La enzima PcCS caracterizada de doble funcionalidad, se reportó que metila 7-metilxantina para formar teobromina y teobromina para formar cafeína en *Paullinia* (Schimpl *et al.* 2013). Sin embargo, la enzima responsable de la biosíntesis de 7- metilxantina se desconoce y la vía biosintética de cafeína en *Paullinia* aún no estaba del todo clara. Mediante el uso de secuencias SABATH de *Citrus* para BLAST contra la base de datos EST de *Paullinia* en Genbank, se ensamblaron cinco secuencias de cafeína sintasa de longitud completa en contigs y se denominaron PaulliniaCS0 a PaulliniaCS4, y estas EST se muestrearon de frutos o semillas de *Paullinia*, donde se informa que la cafeína se acumula (Figueiredo *et al.* 2011; Schimpl *et al.* 2013).

No se encuentra ningún gen de tipo XMT en transcriptoma de *Paullinia*. El análisis filogenético de estas secuencias similares a la cafeína sintasa de *Paullinia* muestra que todas ellas están estrechamente relacionadas con los genes de la cafeína sintasa en *Camellia*, *Theobroma* e *Ilex*. La enzima PaulliniaCS1 fue clonada de *Paullinia* sin ninguna caracterización funcional (Figueiredo *et al.* 2011). PaulliniaCS3 es una variante menor de PaulliniaCS1, y PaulliniaCS4 es una variante menor de PaulliniaCS2. PaulliniaCS1 y PaulliniaCS3 expresados de forma heteróloga mostraron actividades relativas similares y lo mismo para PaulliniaCS2 y PaulliniaCS4. No se ha detectado hasta el momento ninguna actividad *in vitro* de PaulliniaCS0 con los diferentes sustratos de alcaloides de xantina. Se reportó que los análisis de LC-MS de los productos de reacción muestran que PaulliniaCS1 metila la xantina en 3-metilxantina y PaulliniaCS2 metila la 3-metilxantina para formar teobromina. Tanto PaulliniaCS1 como PaulliniaCS2 tienen afinidades de sustrato comparables a las cafeínas sintasas caracterizadas

que están involucradas en la vía biosintética de cafeína en *Coffea* y *Camellia* como, por ejemplo, XMTs en *Coffea* (Uefuji *et al.* 2003) y TCS1 en *Camellia* (Kato *et al.* 1999).

Aunque nunca se informó que la xantina, la xantosina o las mono metilxantinas se acumularon en tejidos de *Paullinia*, se informaron teobromina y teofilina en las semillas de *Paullinia*, donde se acumula la mayor parte de la cafeína (Schimpl *et al.* 2014; Baumann *et al.* 1995; Duke *et al.* 1992). Las actividades relativas de PaulliniaCS1, PaulliniaCS2 y PcCS, junto con los metabolitos encontrados en las semillas de *Paullinia* sugirieron que la producción de cafeína puede haber desarrollado una nueva vía biosintética, que es diferente de las vías en *Camellia*, *Coffea* y *Citrus*: Primero, PaulliniaCS1 metila xantina para formar 3-metilxantina. En segundo lugar, PaulliniaCS2 metila 3-metilxantina para formar teobromina. Tercero, la teobromina se metila para formar cafeína por PcCS (Fig.4).

FIG.4: Biosíntesis de cafeína en *Paullinia cupana*. La síntesis de cafeína comienza a partir de enzimas codificadas por genes tipo CS caracterizadas a partir semillas de *Paullinia cupana*. La vía comienza con la xantina, y los intermediarios son 3-metilxantina y teobromina, diferentes de las encontradas en vías en *Camellia*, *Coffea* y *Citrus*. Estas tres enzimas distintas PcCS1=PaulliniaCS1, PcCS2=PaulliniaCS2 y PcCS son responsables de cada paso de metilación en la ruta. **Fuente:** Adaptado de Huang, R. "Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants" (2017).

1.1.5. BIOSÍNTESIS DE CAFEÍNA EN *Citrus sinensis*(NARANJA)

El principal alcaloide de purina en la flor de los cítricos (principalmente en las anteras y el polen) es la cafeína, seguida de una cantidad significativa de teofilina, así como cantidades insignificantes de teobromina y paraxantina (Kretschmar *et al.* 1999).

Kretschmar y Baumann (1999) también tomaron muestras del contenido de alcaloides de purina en flores de cítricos en diferentes etapas de desarrollo y descubrieron que solo la teofilina es detectable en brotes pequeños. Cuando las flores crecen desde los brotes hasta la antesis completa, una disminución en el nivel de teofilina se acompaña de un aumento en el

nivel de cafeína, mientras que no ocurren cambios en los niveles de teobromina y paraxantina. Estos hallazgos indican que la teofilina es un precursor de la vía para la cafeína, en lugar de la teobromina, y es consistente con las actividades enzimáticas de CitrusXMT2 y la vía caracterizada en *Citrus*. La vía biosintética de la cafeína a través de la teofilina también cambia la visión a largo plazo de que la teofilina es un producto de degradación de la cafeína, como se encuentra en *Camellia*, *Coffea* y *Theobroma* usando trazadores radiomarcados (Zheng *et al.* 2004; Ashihara *et al.* 1996 y 1997).

Además, de las dos vías biosintéticas de cafeína en *Citrus*, la vía a través de 1-metilxantina parece ser una ruta principal a la cafeína, como lo sugieren los parámetros cinéticos de CitrusXMT1 que se obtuvieron con 1- y 3-metilxantina, así como también un nivel alto encontrado en 1-metilxantina más que 3-metilxantina en botones florales de éstos cítricos. Aunque la vía biosintética de la cafeína a través de la xantina a 3-metilxantina a teofilina también se encontró como una vía menor en las hojas jóvenes de las plantines de té (Deng *et al.* 2010), la vía a través de la xantina a 1-metilxantina a la teofilina nunca se había informado en ninguna planta productora de cafeína anteriormente. Sin embargo, *Citrus* parece haber desarrollado la capacidad de sintetizar cafeína a través de esta importante y novedosa vía.

Además de CitrusXMT1 y CitrusXMT2, que están filogenéticamente cerca de los genes XMT en *Coffea*, *Citrus* también posee un gen CS que está estrechamente relacionado con los genes CS en *Camellia*, *Paullinia* e *Ilex*, aunque no tiene actividad detectable con los alcaloides de xantina probados. Su papel en la biosíntesis de cafeína en flores de cítricos es misterioso, porque la vía biosintética de cafeína completa ya está catalizada por los XMT (Fig. 5). Además, los genes CS también se encuentran en algunas plantas que no se sabe que producen cafeína como, por ejemplo, *Kirkia*, *Mangifera*, *Eucalipto*. La caracterización funcional de estos genes CS en el futuro puede ayudar a comprender el papel de CitrusCS en las flores de cítricos.

FIG.5: Biosíntesis de cafeína en *Citrus sinensis*. La cafeína se sintetiza utilizando genes del linaje XMT a través de dos nuevas vías en las flores de cítricos. Las vías comienzan con la xantina, en lugar de la xantosina, y los intermediarios son 1-metilxantina, 3-metilxantina y teofilina, completamente diferentes de la vía propuestas en *Camellia* y *Coffea*. Dos enzimas, CisXMT1=CitrusXMT1 y CisXMT2= CitrusXMT2 son responsables de los tres pasos de metilación y CitrusXMT1 es una enzima multifuncional que cataliza los dos primeros pasos de metilación en cada vía. **Fuente:** Adaptado de Huang, R. "Evolution of Caffeine Biosynthetic Enzymes and Pathways in Flowering Plants" (2017).

1.1.6. CAFEÍNA SINTASA DE *Ilex paraguayensis*

De acuerdo con los datos obtenidos a partir del transcriptoma de la yerba por Debat *et al.* 2014, se pudo identificar solamente una posible secuencia y estructura de la cafeína sintasa que hipotéticamente cataliza los dos últimos pasos de reacción desde la 7-metilxantina a la cafeína, vía teobromina. La hipótesis de la vía biosintética propuesta implicaba reacciones que iban desde la conversión de xantosina a 7-metilxantosina, de 7-metilxantina a teobromina que se convierte en cafeína.

Utilizando una secuencia de CS de *Camellia sinensis* investigaron dentro de su biblioteca e identificaron una secuencia correspondiente a la longitud total de un transcrito del transcriptoma de la yerba mate. La hipotética CS de yerba mate se había identificado por similitud con el ARNm completo de CS de *C. sinensis*.

De acuerdo a ésta predicción, la secuencia abarcaba unos 1.098 nt que codifica una proteína 366 aa con una identidad del 61% a la correspondiente CS de *C. sinensis* de 369 aa, compartiendo el dominio de la metiltransferasa_7 SAM dependiente de carboxil metiltransferasa implicado en la síntesis de cafeína. Comparando las estructuras de la proteína de *Coffea arabica*, *Camellia sinensis* y *Theobroma cacao* con el transcrito predicho de yerba

mate pudieron observar conservación génica y también del dominio y, asimismo, generar un modelo 3D de la estructura cuaternaria de la CS predicha de yerba mate con el café (Fig. 6).

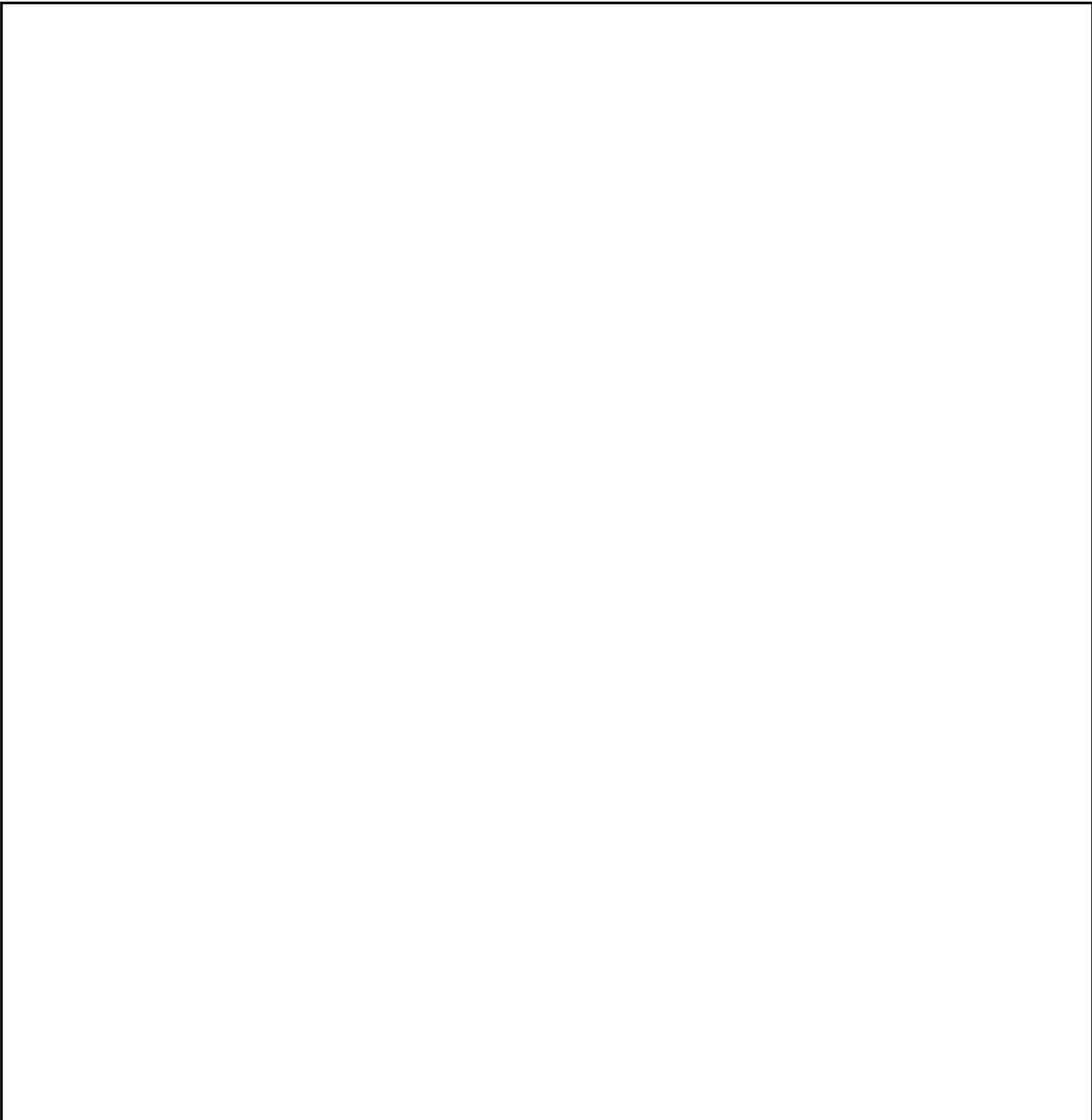


FIG.6: Estructura 3D de la cafeína sintasa en *Ilex paraguariensis*. (CS). Empleando la estructura resuelta de cristalografía de rayos X de CS de *Coffea arabica* como molde (b), la estructura 3D de la yerba mate CS fue predicha por el algoritmo del SWISS-model (a). Un modelo de cinta de CS de yerba (c) y CS de café (e) sugiere una alta conservación de la estructura secundaria cuando se superponen (d). Se presenta una reconstrucción de un modelo de malla de yerba CS (f) y se compara con el café EM (h), que muestra una gran similitud de estructura cuaternaria (g). **Fuente:** Debat *et al.* 2014.

1.2. HIPÓTESIS Y OBJETIVOS

Tesis Doctoral

Sebastián Maximiliano Rossi

La vía de biosíntesis de la cafeína es una vía común a una serie de especies de gran importancia económica por la generación de teobromina y cafeína como bioproductos de gran impacto económico y cultural. Sin embargo, el balance de estos productos en el funcionamiento natural de la vía en cada especie está influenciado por diversos parámetros genéticos y ambientales. El conocimiento de la genómica comparativa de las enzimas involucradas en esta vía permitirá cuantificar el peso de las variaciones genéticas en este balance y seleccionar marcadores predictivos del mismo.

Tomando como base los fundamentos relacionados al impacto e importancia de la cafeína, el objetivo del trabajo fue analizar y describir desde el punto de vista genómico y filogenético la vía biosintética de la cafeína de la YM comparativamente con la existente en otros organismos de importancia agronómica. De esta forma, poder predecir el impacto de las variaciones que se puedan encontrar a nivel proteómico y metabólico, y al mismo tiempo caracterizar todas las enzimas involucradas en dicha vía biosintética comparándola con otras especies productoras de cafeína de interés contrastando con especies no productoras. De esta manera se pretende mostrar cambios únicos y también comprender la evolución de rasgos en bloques génicos u operones conservados en el desarrollo de algunas angiospermas en especies que realizan la misma función y denotar la falta de estos grupos génicos en especies que no producen cafeína mediante estudios comparativos de sintenia y análisis de genes ortólogos.

1.2.1. OBJETIVOS ESPECÍFICOS

- 1.** Analizar metabólicamente la vía de la cafeína en YM y seleccionar las enzimas de interés.
- 2.** Caracterización de las enzimas principales y secundarias participantes de la vía metabólica de la YM.
- 3.** Comparar las secuencias génicas correspondientes a las enzimas de interés entre especies emparentadas y entre especies de interés productivo.
- 4.** Análisis evolutivo de las enzimas primarias y secundarias de la vía metabólica de la YM con especies productoras de cafeína de interés comercial y de las no productoras.

OBJETIVO	HITO DE EVALUACIÓN
1. Análisis metabólico de la vía de la cafeína y selección de enzimas de interés.	Identificar los principales puntos de control de la vía metabólica
2. Caracterización de las enzimas principales y secundarias participantes de la vía metabólica	Destacar los aminoácidos clave involucrados en los sitios activos de las proteínas mediante un estudio molecular de <i>Docking</i>
3. Comparar las secuencias génicas correspondientes a las enzimas de interés entre especies emparentadas y entre especies de interés productivo.	Emparejamiento de secuencias, detección de regiones polimórficas, de genes ortólogos y parálogos
4. Análisis evolutivo de las enzimas primarias y secundarias de la vía metabólica con especies productoras de cafeína de interés comercial y de las no productoras.	Mediante un estudio de microsintenia de la región de genes de cafeína visualizar como el bloque genómico evoluciona en especies comparadas destacando eventos de duplicación y fraccionamiento génico.

2. MARCO TEÓRICO

La yerba mate es ampliamente cultivada en Argentina, Brasil y Paraguay en un área conjunta de 330.000 hectáreas (Jerke *et al.* 2009; INYM 2014). Argentina es el mayor productor global, con un rendimiento de más de 1.000.000 de toneladas anuales que representan alrededor de 80% del total a nivel internacional (INYM 2014). La provincia de Misiones está a la vanguardia en el país, abarcando la mayor superficie cultivada (90%) y concentrando el 97% de los productores. Asimismo, Misiones concentra la mayor parte de la etapa industrial con el mayor número de secaderos (94%), acopiadores (93%) y molinos (85%). Por ello, esta provincia genera el 84% de la producción nacional, seguida por la provincia de Corrientes que contribuye con el 16% restante (Canitrot *et al.* 2011; INYM 2014). Particularmente en Misiones, el cultivo ocupa el 50% de su superficie agrícola y la actividad yerbatera en esta provincia representa el 22% de la mano de obra en el sector primario y el 15% en el sector manufacturero y constituye la primera actividad industrial con el 25% del total (Canitrot *et al.* 2011; Dolce 2012) lo cual enfatiza su importancia para las economías locales. Particularmente, desde la creación del Instituto Nacional de la yerba mate (INYM) en 2002, se observa un significativo aumento de la participación relativa de los productores y de los secaderos en el precio final de la yerba (Gortari 2007; Rau *et al.* 2009). Alrededor del 15% de la producción total de yerba mate se exporta a mercados de Sudamérica, Europa y Asia, destacándose la participación de Siria (Canitrot *et al.* 2011). Sumada a su importancia agrícola y económica, es de destacar la profunda y marcada influencia de la yerba mate en la dinámica sociocultural de América Latina y especialmente en el Mercosur.

Su amplia versatilidad en la industria ha suscitado el estudio de especies co-genéricas, en su mayoría simpátricas, consideradas como adulterantes (Filip *et al.* 1998; Barchuk *et al.* 2013). Adulterantes de otras especies de *Ilex* pueden ser incorporados en el producto final, intencionalmente o no. Se investigaron en seis especies de *Ilex* encontradas como adulterantes de té de yerba las cantidades de teobromina, teofilina y cafeína. Las especies

analizadas fueron *I. dumosa*, *I. pseudobuxus*, *I. brevicuspis*, *I. theezans*, *I. microdonta*, e *I. argentina*. Como resultado se observó que estas especies contenían pequeñas cantidades o ninguna de los compuestos investigados. Sólo se encontraron trazas de cafeína en *I. theezans*, *I. dumosa*, *I. microdonta* e *I. pseudobuxus*. Sólo trazas de teobromina se encontraron en *I. argentina* e *I. microdonta*. La teofilina fue sólo cuantificable en *I. pseudobuxus* a 6 ppm (Filip *et al.* 1998). Utilizando HPLC y NMR para analizar cafeína y teobromina de variedades de *Ilex*, se encontraron estas xantinas sólo en *Ilex paraguariensis* al compararse con sus adulterantes (Reginatto *et al.* 1999; Choi *et al.* 2005).

Las especies de *Ilex* adulterantes pueden ser problemáticas en la calidad del producto debido a sus diferencias en la concentración de saponinas. El té de yerba preparado con la planta de *Ilex paraguariensis* demostró ser menos amargo que todos los extractos preparados con las especies adulterantes. Por lo tanto, es posible que la adición de estas especies pueda tener un efecto significativo en el sabor amargo de las bebidas hechas con yerba mate. No sólo las plantas adulterantes pueden tener grandes concentraciones de compuestos amargos, sino que los frutos de la planta de *Ilex paraguariensis* también contienen saponinas amargas. Si estos frutos fueron incorporados a los productos del mate, esto incrementaría el sabor amargo disminuyendo su calidad (Taketa, 2004).

De todas las especies de *Ilex*, *I. dumosa* es el adulterante más importante y con saponinas de estructura similar a *Ilex paraguariensis*. Debido a la especificidad de las saponinas, es posible identificar adulterantes de yerba mate basándose en la concentración de estas sustancias (Pires *et al.* 1997).

Los estudios sobre las propiedades medicinales de la yerba mate se han incrementado en la última década y existe un consenso actual sobre sus efectos benéficos como antioxidante, anti-inflamatorio, antimutagénico y sobre el metabolismo lipídico (Filip *et al.* 2000, 2001; Anesini *et al.* 2006, 2012; Bracesco *et al.* 2011) que renuevan el interés en el estudio de esta importante especie.

2.1. GENÉTICA DE LA YERBA MATE

Se han reportado marcadores cromosómicos de tipo RFLP y AFLP (Gauer *et al.* 2000; Gottlieb *et al.* 2011; Stein *et al.* 2014). En relación al conocimiento del genoma de yerba mate, Debat

et al. 2014, han publicado por primera vez un borrador completo de su transcriptoma, del cual se deduce la presencia de al menos 32.000 genes y 12.000 variantes. Este trabajo explora la vasta colección de transcritos de *I. paraguariensis* utilizando como genoma de referencia a la línea Pg538, material elite desarrollado por el INTA EAA-Cerro Azul de Misiones. El ARN total de *I. paraguariensis* fue secuenciado con la plataforma Illumina HiSeqTM-2000 obteniéndose 72.031.388 secuencias de 100 pb con una cobertura estimada de 180X.

En líneas generales, se identificaron y categorizaron transcritos pertenecientes a más de 100 vías metabólicas. Asimismo, como se ha comentado en el capítulo 1 sección de antecedentes, se identificaron en este estudio alrededor de 1.000 factores de transcripción putativos, genes implicados en estrés por calor y estrés oxidativo, respuesta a patógenos, resistencia a enfermedades y respuesta a hormonas. También se identificaron transcritos relacionadas a estrés osmótico, sequía, salinidad, estrés por frío, senescencia, floración temprana, y determinación sexual.

La gran cantidad de información obtenida en este estudio sirvió de marco de referencia y dio lugar al inicio del “Proyecto de Secuenciación del Genoma de Yerba Mate” (Pro.Mate.A.R.) en 2015 por parte de la Universidad Nacional de Misiones (UNaM), la Universidad de Buenos Aires (UBA) y otras instituciones argentinas (Secretaría de políticas Universitarias, Instituto Nacional de la Yerba Mate, CONICET, Ministerio de Ciencia, Tecnología e Innovación Productiva) que tiene por objetivo generar un mapa físico completo a nivel de secuencia de todo el genoma de *I. paraguariensis* *St. Hilarie*. Esto constituye el punto de partida para próximos estudios tales como: descubrir, caracterizar y localizar genes asociados a caracteres de importancia biológica, agronómica y económica; desarrollar marcadores moleculares eficientes que puedan emplearse para identificar dichos genes mediante mapeo genético por asociación o en progenies segregantes; analizar la diversidad genética y realizar cruzamientos selectivos.

El conocimiento generado sobre la genómica y la transcriptómica de *I. paraguariensis* permitirá comprender con profundidad su bioquímica y evaluar el impacto de las variantes alélicas que se presenten sobre la metabolómica de la especie generando la capacidad de implementar mejores técnicas de selección de individuos que muestran un balance metabólico determinado, como por ejemplo una mayor o menor producción de cafeína, basándose en las características proteómicas de las enzimas que forman parte de la vía. Esta posibilidad podrá impactar no solo en el conocimiento de las vías de producción de

fitoquímicos y productos con interés farmacológico, que permita compararlas con las existentes en otras plantas modelo, sino cambiar la base de selección de individuos destacados basándose en la búsqueda de variantes alélicas.

2.2. IMPORTANCIA AGRONÓMICA

Ilex paraguariensis St. Hil. es un árbol dioico del orden de los Aquifoliales, familia Aquifoliaceae del grupo de las Euastéridas II (Fig. 7) originario de las selvas subtropicales de América del Sur. Sus hojas y pequeñas ramas secas procesadas son utilizadas para preparar la infusión popular, estimulante y de sabor particular, conocida como “mate”, de allí que su nombre coloquial en América Latina sea “yerba mate” o “erva mate”, expresión derivada de la conjunción “ka`a mati”, del guaraní “ka`a” (yerba) y el quechua “mati” (calabaza), esta última en alusión al recipiente de consumo. La yerba mate también se utiliza ampliamente para preparar otras infusiones y brebajes calientes como el “mate cocido”, bebidas frías como el “tereré” y más recientemente, como aditivo en helados, dulces y bebidas energéticas, así como en tintes, cosméticos e ingredientes de spa (Bracesco *et al.* 2011).

Los cultivos de yerba mate son un importante agronegocio en Misiones, Argentina. Las características y condiciones agroecológicas de los cultivos de yerba mate son encontrados solamente en Brasil, Paraguay y Argentina. En Argentina, el primer productor mundial, las plantaciones se limitan a las provincias de Misiones y Corrientes.

El relevamiento aerofotogramétrico de plantaciones encargado por el INYM (Instituto Nacional de la Yerba Mate, ente nacional que regula todos los aspectos referidos al cultivo y la comercialización del producto) en el año 2016, determinó que el cultivo de yerba mate en Argentina abarca una superficie de 165.200 hectáreas, de las cuales 144.014 hectáreas están localizadas en la provincia de Misiones (87%) y las demás se encuentran en la provincia de Corrientes.

FIG. 7: Cladograma en el cual se expresan las interrelaciones entre órdenes y familias de Angiospermas aportadas por el análisis de Jackknife con frecuencias superiores a 50% (APG II, 2003 De acuerdo al APG II (2003) y a Stevens (2001) los principales grupos, formales y no formales, de las Angiospermas son los que siguen a continuación. En el recuadro rojo se indica el orden a la cual pertenece *Ilex paraguariensis* dentro del grupo de las euastéridas II.

En Argentina se consume yerba mate en forma habitual en el 98% de los hogares (76% de la población de este país), estando el 63% de este consumo destinado a paquetes de yerba mate para la infusión tradicional, con un promedio de consumo de 2,8 Kg mensuales. Durante el 2013 el consumo de yerba mate se incrementó a 257.000 tn; 6,44 kg per cápita, representando el 87% de la producción de yerba mate de ese año. Esto destaca la alta dependencia que tiene este cultivo del consumo interno ya que representa hasta el 50% del consumo de bebidas de la población.

Este crecimiento del consumo trae asociada la necesidad de optimizar el rendimiento de los cultivos para incrementar la producción y dar respuesta al crecimiento del mercado de yerba mate. En contrapartida al crecimiento en el consumo, y de acuerdo a informes del INYM, en los últimos años se ha observado una caída sustancial en la producción que podría estar asociada a una disminución de la fertilidad de los campos (manejo destructivo de los suelos), longevidad de las plantas, material de baja calidad genética, prácticas agrícolas inadecuadas en tiempo, podas incorrectas, entre otros (Prat Kricun, 2003).

La necesidad de incrementar el rendimiento de los campos y el interés de los estamentos gubernamentales y sectoriales hacia las prácticas amigables y sustentables representa una gran oportunidad para la incorporación nuevos conocimientos que permitan mejorar el rendimiento de los cultivos (Bergottini *et al.*, 2015; López *et al.*, 2019) y la obtención de productos alternativos que fortalezcan la cadena de valor y generar mercados alternativos. Estudios sobre las propiedades físicoquímicas y medicinales de la yerba mate se han incrementado en la última década y existe un consenso actual sobre sus efectos benéficos como antioxidante, antiinflamatorio, antimicrobiano y regulador del metabolismo lipídico (Filip *et al.* 2000, 2001; Anesini *et al.* 2006, 2012; Bracesco *et al.* 2011) que incrementan el interés en fitomedicina y biotecnología de esta importante especie.

En concordancia con esto, los lineamientos del “Plan Estratégico para el Sector Yerbatero Argentino 2013-2028” proponen el mejoramiento genético de la yerba mate como una de las claves para el ascenso de las economías locales de la provincia de Misiones, íntimamente relacionadas con el cultivo, las cuales se beneficiarían con el valor agregado que supondría la introducción de cultivares de origen certificado y *blend*, con mejores características organolépticas relacionadas con el sabor (asociado con una variación en los contenidos de cafeína, teobromina, glucosa y sacarosa) y con el aroma (asociado con la presencia de más de 40 compuestos volátiles). Asimismo, los productores se beneficiarían con la introducción de

plantas con una mayor uniformidad en la densidad de ramas, tiempo de brotación, rendimiento unitario y tolerancia a plagas específicas de la yerba mate como el psílido, taladro grande, marandová y también con variedades adaptadas a diferentes regiones y exigencias como la tolerancia a períodos de estrés hídrico, bajas temperaturas, heladas y alta insolación.

2.3. IMPORTANCIA DE LA VÍA DE BIOSÍNTESIS DE CAFEÍNA

Un aspecto destacado del proyecto es la comprensión de la vía metabólica a través de la cual se generan los bioproductos de alto valor agregado como son los alcaloides que le confieren a la planta muchas de las propiedades de interés nutricional, biomédico y bioquímico. Estos conocimientos son de relevancia para evaluar las potencialidades en la selección de variantes con diferentes contenidos en alcaloides destinados a la producción de bebidas energizantes, nuevas formulaciones de yerba mate de gran interés alimentario y nutricional, productos alimenticios combinados, productos cosméticos y farmacéuticos.

2.4. PRODUCCIÓN DE CAFEÍNA EN ANGIOSPERMAS

La evolución convergente es un proceso que ha ocurrido a lo largo del árbol de la vida, pero el contexto histórico genético y bioquímico que promueve independientes orígenes repetidos de un rasgo es rara vez entendido. La conocida cafeína y sus precursores alcaloides de xantina, han evolucionado varias veces en la historia de plantas con flores para varios roles como mecanismos de defensa y también para la polinización de estas plantas. Se ha demostrado que la producción convergente de cafeína, sorprendentemente, ha evolucionado por dos vías bioquímicas previamente desconocidas en plantas de cacao, cítricos y guaraná que usan cafeína sintasa (CS) o enzimas tipo metiltransferasa de xantina (XMT). Sin embargo, el camino y el linaje enzimático utilizado por cualquier especie de planta dada no es predecible por una relación filogenética sola. La resurrección de secuencias ancestrales revela que esta convergencia fue facilitada por la co-optación de genes mantenidos por más de 100 millones de años para roles bioquímicos alternativos. Las antiguas enzimas del linaje de los cítricos fueron exaptadas para reacciones actualmente utilizadas para varios pasos de la biosíntesis de cafeína y requirió muy pocas mutaciones para adquirir características enzimáticas modernas, lo que permite la evolución de una vía completa. Los estudios futuros destinados

a manipular el contenido de cafeína de las plantas requieren el uso de diferentes enfoques dado el metabolismo y diversidad genética que poseen los distintos órdenes de plantas productoras de cafeína (Huang *et al.* 2016).

2.5. GENES CONSERVADOS EN LA EVOLUCIÓN

Muchas funciones genéticas nuevas surgieron a través del tiempo de evolución de las especies y ha sido un área de investigación importante. Una de las propuestas del surgimiento de dichas funciones novedosas fueron la duplicación génica en el cual se produce un aumento de genes que se encuentran bajo selección natural. La duplicación de genes ha sido reportada en todos los dominios de la vida. Nuevas copias de genes pueden ser generadas por eventos de recombinación de ADN como por ejemplo *crossing-over* desigual o retroposición mediada por ARN o por duplicación completa del genoma (Whole-genome Duplication WGD). Se ha documentado que los genes duplicados tienen varios destinos evolutivos.

Muchos genes se mantienen mediante la provisión de una cantidad adicional de producto genético beneficioso (dosis aumentada), pueden deberse también a que adquieren funciones novedosas o heredan funciones ancestrales parciales que son complementarias a las funciones de la otra copia génica o los genes se convierten en pseudogenes al acumular mutaciones perjudiciales y eventualmente se pierden del genoma.

Se ha argumentado que la evolución del tamaño del genoma de la planta es principalmente unidireccional y aumenta debido a la acción variada de las duplicaciones completas del genoma (WGD) y la proliferación de elementos móviles. Una interacción de las tasas de delección y proliferación de retroelementos se relaciona con una continua evolución del tamaño del genoma de la planta, con eventos de WGD que proporciona amortiguación a corto plazo contra la pérdida de funciones genéticas cruciales en genomas pequeños afectados por altas tasas de delección endógena. Los genomas pequeños se producen cuando la tasa de delección recombinacional es alta en relación con la proliferación de retroelementos y WGD, viceversa con genomas grandes. (Ibarra-Laclette *et al.* 2013).

Muchas sustancias químicas se producen a través de vías metabólicas que se componen de una serie de reacciones catalizadas por enzimas, en lugar de reacciones simples de un solo paso. Sin embargo, a pesar muchos de estudios de estas vías, solo hay unas pocas investigaciones de cómo se ensamblaron evolutivamente las vías metabólicas que presentan varios pasos; en otras palabras, en qué orden evolucionaron las enzimas individuales para cada paso.

Durante este trabajo doctoral se llevó a cabo el análisis evolutivo y filogenético de la manera en que evolucionó la vía biosintética de la cafeína en yerba mate comparándola con grupos de plantas productoras de cafeína. Por tal motivo se tomaron en consideración tres hipótesis para explicar el ensamblaje de las rutas metabólicas.

La primera hipótesis con respecto al origen de las rutas metabólicas fue la **hipótesis retrógrada** (Horowitz, 1945), que establece que el ensamblaje de una ruta metabólica se produjo en orden inverso a la dirección de la biosíntesis de manera gradual. Específicamente, el último paso y su enzima asociada en la vía evolucionaron primero, luego la duplicación de genes y la selección natural facilitaron la evolución de una segunda enzima que catalizó el penúltimo paso, y el proceso continuó hasta que finalmente se ensambló el primer paso. Horowitz mencionó que se necesita un trasfondo ambiental especial para esta hipótesis: un ambiente primitivo con intermedios de vías producidos de forma no enzimática, ya que las células no habrían tenido la capacidad de sintetizarlos directamente.

En otras palabras, ésta idea tenía un atractivo intuitivo y establecía que, si la biosíntesis contemporánea del compuesto "A" requiere la transformación secuencial de los precursores "D", "C" y "B" a través de las enzimas correspondientes, el final del producto "A" de una ruta metabólica dada fue el primer compuesto utilizado por los heterótrofos primitivos (Fig. 8). Entonces, si el compuesto "A" era esencial para la supervivencia de las células primitivas, cuando "A" se agotaba del caldo primitivo, esto debería haber impuesto una presión selectiva que permitiera la supervivencia y reproducción de aquellas células que habían podido funcionar la transformación de un compuesto químicamente relacionado "B" en "A" catalizado por la enzima "a" que habría llevado a una ruta simple de un solo paso. La selección de variantes que tienen una enzima "b" mutante relacionada con "a" a través de un evento de

duplicación y capaz de mediar la transformación de la molécula "C" químicamente relacionada con "B" conduciría a una ruta cada vez más compleja, un proceso que continuaría hasta que toda la ruta se estableciera hacia atrás, comenzando con la síntesis del producto final, luego la penúltima ruta intermedia, y así sucesivamente hasta el precursor inicial. Muchos años después de la propuesta de esta hipótesis, el descubrimiento de operones llevó a Horowitz a reformular su modelo, argumentando que también estaba respaldado por los clústeres de genes, lo que podría explicarse por una serie de duplicaciones tempranas en tándem de un gen ancestral; de este modo, los genes que pertenecen al mismo operón o a la misma vía metabólica deberían haber formado una familia de genes parálogos.

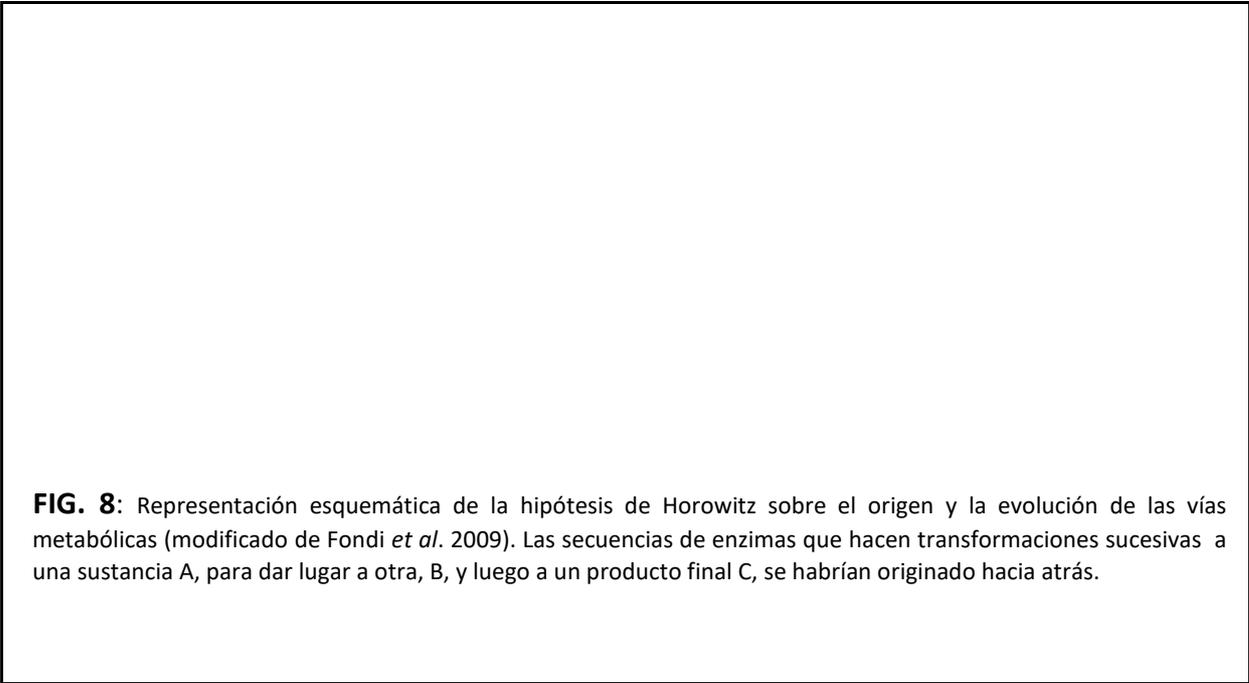


FIG. 8: Representación esquemática de la hipótesis de Horowitz sobre el origen y la evolución de las vías metabólicas (modificado de Fondi *et al.* 2009). Las secuencias de enzimas que hacen transformaciones sucesivas a una sustancia A, para dar lugar a otra, B, y luego a un producto final C, se habrían originado hacia atrás.

Esta hipótesis tiene su limitación porque es poco probable que exista un entorno de este tipo, y no considera si los intermedios en una vía eran estables o no y cómo se transportaron estos productos químicos a través de una red celular, incluso si estuvieran disponibles.

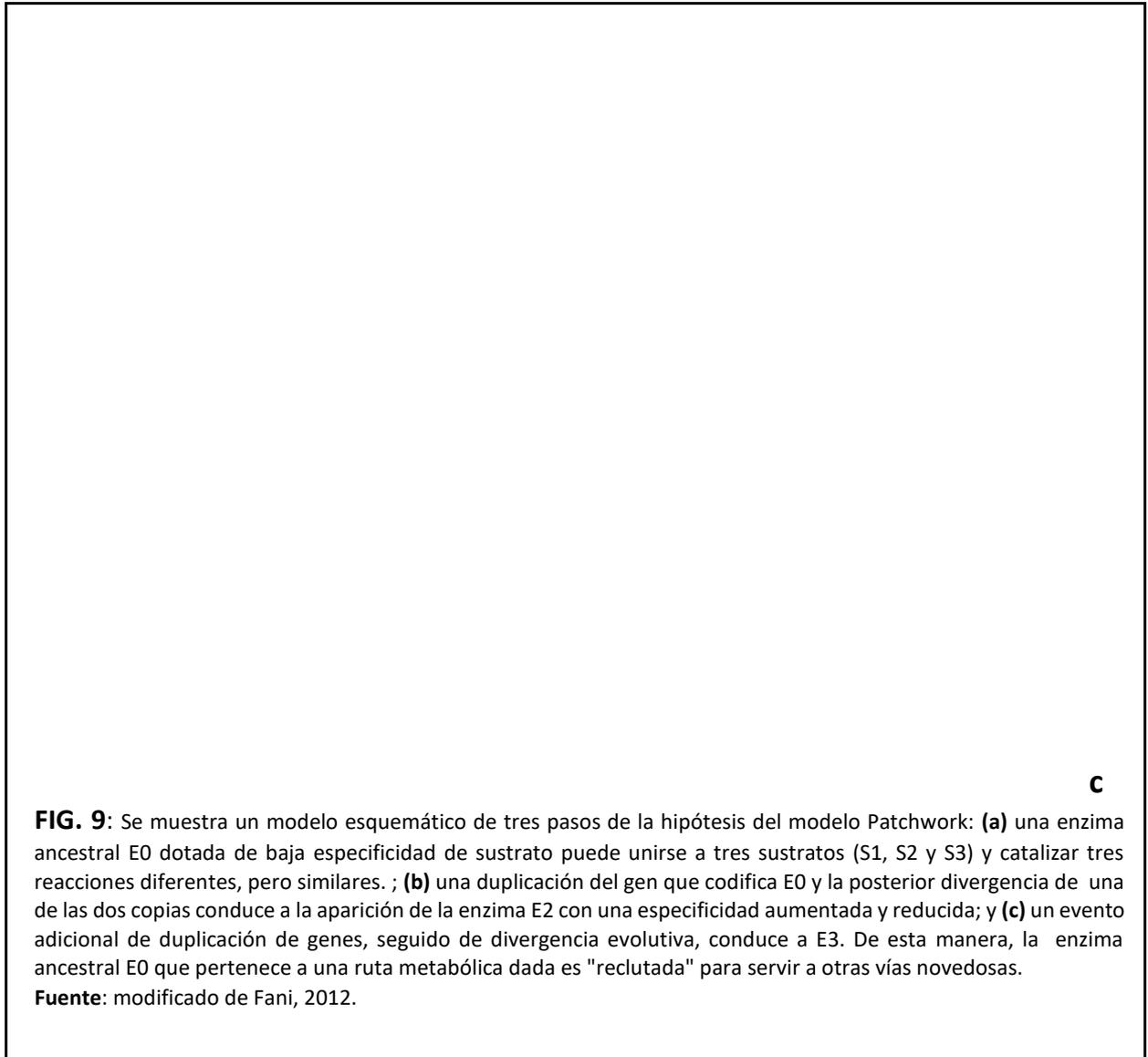
La segunda hipótesis se denominó **hipótesis acumulativa** (o hipótesis de Granick), que era contraria a la hipótesis retrógrada. Con evidencia de la evolución de la ruta fotosintética de una unidad primitiva de conversión de energía, Granick propuso que el ensamblaje de una ruta metabólica se realizará en la misma dirección que la síntesis del producto de manera gradual. Específicamente, las reacciones anteriores y las enzimas asociadas en una vía

evolucionaron primero y comenzaron a partir de compuestos de estructura simple. Posteriormente, los pasos segundo y tercero evolucionaron secuencialmente con una complejidad creciente en las estructuras de los productos intermedios y finales, así como el reclutamiento de enzimas más eficientes a través de duplicaciones de genes. Por ejemplo, según Granick, al ser la biosíntesis de clorofilas un paso intermedio en la biosíntesis de bacterioclorofilas, los organismos que realizan fotosíntesis oxigénica y contienen clorofilas precedieron a los que contienen bacterioclorofilas y realizan fotosíntesis anoxigénica. Sin embargo, la filogenia molecular demuestra que la bacterioclorofila *a* es más ancestral, lo que implica que los centros de reacción cianobacterianos que contienen clorofila *a* son evolutivamente más recientes y que la evolución de bacterioclorofilas a clorofilas supuso pérdida de genes y acortamiento de la ruta de biosíntesis.

Aunque la hipótesis acumulativa se aplicó con éxito para explicar la evolución del colesterol de los terpenoides, tiene su limitación ya que los intermediarios en una vía deben ser útiles para el organismo, porque el reclutamiento de enzimas multipaso en una vía rara vez ocurriría simultáneamente.

A diferencia de las dos primeras hipótesis, la **hipótesis de evolución en retazo** o modelo *patchwork* de los bioquímicos Waley, Jenssen e Ycas desarrollado respectivamente en 1969, 1974 y 1976, no se centra en el orden de los pasos desarrollados (o las enzimas reclutadas) en una vía. En cambio, se centra en la promiscuidad del sustrato de las enzimas ancestrales y el ensamblaje de nuevas vías metabólicas mediante el reclutamiento de enzimas promiscuas que catalizan de manera fortuita las reacciones existentes. La importancia clave de la hipótesis de evolución en retazo es que las actividades promiscuas de las enzimas ancestrales permitieron el desarrollo instantáneo de una nueva vía al hacer uso de estas actividades menores. Una vez que el producto de la nueva vía se vio favorecido por la selección natural, las duplicaciones de genes y las mutaciones funcionales de enzimas ancestrales proporcionaron enzimas recientemente especificadas para completar la nueva vía. Esta hipótesis tuvo en cuenta la naturaleza generalizada de los eventos de duplicación de genes y las actividades promiscuas comúnmente observadas de las enzimas, por lo que tiene una aplicabilidad mucho más amplia que cualquiera de las dos primeras hipótesis, especialmente para la evolución de las vías metabólicas secundarias de la planta (Fig. 9).

En este modelo *patchwork* tales enzimas inespecíficas o promiscuas eran relativamente lentas y pueden haber permitido que las células primitivas que contienen genomas pequeños superen sus capacidades de codificación limitadas.



La hipótesis de retazos también es consistente con la posibilidad de que una vía ancestral pueda haber tenido una enzima primitiva que cataliza dos o más reacciones similares en sustratos relacionados de la misma ruta metabólica y cuya especificidad de sustrato se refinó como resultado de eventos de duplicación posteriores. De esta manera, las células primitivas podrían haber expandido sus capacidades metabólicas. Además, este mecanismo puede

haber permitido la evolución de mecanismos reguladores coincidentes con el desarrollo de nuevas vías (Fani, 2004; Lazcano *et al.* 1995).

Mientras que las enzimas que catalizan pasos consecutivos en una ruta metabólica generalmente no son parálogas, lo que favorece la hipótesis de retazos, hay varios ejemplos que muestran que las enzimas codificadas por genes duplicados fueron reclutadas para catalizar pasos consecutivos en la misma ruta metabólica, como se requiere de la hipótesis retrógrada (Fig. 10). Por ejemplo, los genes duplicados *metB* y *metC* que codifican para la cistationina sintasa y β liasa en *E. coli*, respectivamente, evolucionaron para catalizar etapas consecutivas en la biosíntesis de metionina. De manera similar, se ensamblaron cuatro ligasas de carbono-nitrógeno parálogas para catalizar una ruta biosintética de peptidoglucano de cuatro pasos en *E. coli*. Las vías biosintéticas de cafeína en tres pasos caracterizadas en *Camellia*, *Citrus*, *Paullinia* y *Coffea* también son catalizadas por enzimas parálogas de tipo CS o XMT duplicados.

Para probar las tres hipótesis y descubrir cómo se ensamblaron evolutivamente estos tres pasos de metilación en cada especie, es necesario buscar respuestas en la historia evolutiva de dichas plantas.

El desarrollo de técnicas de biología paleomolecular (secuencia de resurrección ancestral junto con mutagénesis experimental) permite responder estas preguntas y desentrañar la historia evolutiva de la biosíntesis de cafeína en las plantas.

2.6.1. RESURRECCIÓN DE SECUENCIAS ENZIMÁTICAS ANCESTRALES

La idea de resurrección de secuencias ancestrales (RSA) se planteó por primera vez en 1963 (Pauling *et al.* 1963). Esta técnica utiliza métodos estadísticos como, por ejemplo, parsimonia, máxima probabilidad o inferencia bayesiana, para estimar la secuencia ancestral de genes o proteínas en un contexto filogenético, para luego sintetizar químicamente el gen y probar directamente su función a través de la expresión del gen heterólogo (Thornton, 2004).



FIG. 10: Comparación esquemática de las tres hipótesis sobre el origen y la evolución de las vías metabólicas en plantas. La hipótesis retrógrada (A) es opuesta a la hipótesis acumulativa (B) en el orden (tiempo) en el que se ensamblaron los pasos. Sin embargo, ambos afirman que el camino se ensambló de manera gradual. La hipótesis en retazos o modelo patchwork (C) se centra en el reclutamiento de enzimas que catalizaron reacciones análogas en una nueva vía al hacer uso de sus actividades promiscuas. La duplicación de genes y la divergencia funcional de estas enzimas proporcionaron la nueva vía con enzimas recientemente especializadas.

La RSA se usó ya en 1990 para estudiar la propiedad de una antigua ribonucleasa de un rumiante bóvido extinto (Stackhouse *et al.* 1990), y más tarde se aplicó gradual y ampliamente en los espectros de absorbancia de pigmentos visuales antiguos (Yokoyama y Shi, 2000), el paleoambiente de bacterias antiguas (Gaucher *et al.* 2003), la evolución funcional de los receptores de corticoides (Bridgham *et al.* 2006) y el desarrollo del color de las proteínas fluorescentes en los corales (Alieva *et al.* 2008).

Estos y varios otros estudios demostraron que la RSA es una herramienta invaluable para comprender la vida antigua, especialmente el comportamiento y la fisiología de los

organismos ancestrales, así como sus interacciones con el entorno antiguo, y proporciona un complemento útil para el registro fósil.

Los resultados de los estudios de RSA muestran que algunas proteínas ancestrales resucitadas eran funcionalmente específicas. Por ejemplo, se descubrió que las primeras esterasas ancestrales que dieron origen a las esterasas modernas y a las acetonas cianohidrina liasas eran muy específicas como las modernas esterasas, que separa los ésteres en un alcohol y un ácido carboxílico a través de un mecanismo de dos pasos, sin embargo, carecen de la capacidad de clivaje de la acetona cianohidrina (Rauwerdink *et al.* 2016). En contraste, la mayoría de los estudios de proteínas ancestrales muestran que eran funcionalmente promiscuas, porque exhibían más de una función o tenían actividades catalíticas con más de un sustrato. Por ejemplo, el receptor corticoide ancestral de los vertebrados podría ser activado por la desoxicorticosterona, la aldosterona y el cortisol (Bridgham *et al.* 2006), la maltasa ancestral pre-duplicada en la levadura tenía actividades con sustratos de maltosas y similares a la isomaltosa (Voordeckers *et al.* 2012), las precámbrica β -lactamasas bacterianas podrían degradar más de un tipo de antibiótico (Risso *et al.* 2013), y las proteínas de unión a aminoácidos polares ancestrales en bacterias fueron capaces de unirse tanto a L-arginina como a L-glutamina (Clifton y Jackson, 2016). Un ejemplo de utilización del método en plantas demostraba que los ancestros resucitados de SAMT en el linaje Apocynaceae y Solanaceae también eran multifuncionales y podían metilar ácido benzoico, ácido salicílico y ácido nicotínico (Huang *et al.* 2012).

Es importante destacar que estas actividades promiscuas tenían la hipótesis de proporcionar el punto de partida para la evolución de nuevas funciones génicas o vías metabólicas. Por lo tanto, estos estudios apoyan la hipótesis de retazos y predicen que el organismo antiguo adquiriría ventajas selectivas en entornos cambiantes debido a las actividades fortuitas promiscuas, y la posterior duplicación o amplificación de genes permitiría la especialización de nuevas funciones génicas a través mutaciones (Ycas, 1974).

Se ha encontrado evidencia considerable para apoyar esta hipótesis, incluida la evolución de

la maltasa ancestral mencionada anteriormente, las β -lactamasas y la proteína de unión a aminoácidos polares, así como los estudios de la evolución de laboratorio dirigida utilizando *E. coli* (Matsumura y Ellington, 2001).

En general, estos estudios indican que las proteínas ancestrales tenían una función primaria que era esencial para el organismo antiguo, mientras que otras actividades podrían no tener ningún papel definido (Fig. 11). Sin embargo, una de estas actividades menores podría haberse seleccionado más adelante para convertirse en la actividad principal de una proteína recientemente desarrollada, un fenómeno denominado *exaptación* (Gould y Vrba, 1982). Por ejemplo, las plumas en pájaros podrían usarse inicialmente para regular la temperatura corporal, pero ha evolucionado para el vuelo más recientemente. Por lo tanto, las plumas fueron exaptadas por la capacidad de volar en aves ancestrales. Del mismo modo, la enzima ancestral multifuncional fue exaptada para actividades menores que más tarde se convirtieron en ventajas selectivas.



FIG. 11: Árbol filogenético que indica de dónde se derivan las secuencias RSA, así como la diferencia entre la comparación horizontal y vertical. (Dr. Lewis Pirenne, 2016)

2.6.2. EVOLUCIÓN CONVERGENTE DE CAFEÍNAS SINTASAS EN RÓSIDAS Y ASTÉRIDAS

Las relaciones filogenéticas de las cafeínas sintasas en *Camellia*, *Paullinia*, *Coffea* y *Citrus* muestran que las cafeínas sintasas dentro de cada especie están más estrechamente

Página | 42

Tesis Doctoral
Sebastián Maximiliano Rossi

relacionadas entre sí que con las de otras especies. Esto indica que se han producido duplicaciones de genes dentro de cada especie después de los eventos de especiación y se espera que dichas secuencias duplicadas se conserven solo si les confieren una ventaja selectiva. Cuando se comparan entre diferentes especies, las cafeínas sintasas en *Camellia* y *Paullinia* están más estrechamente relacionadas (codificadas por genes ortólogos en el clado CS), mientras que la cafeína sintasas en *Citrus* y *Coffea* están más estrechamente relacionadas (codificadas por genes ortólogos en el clado XMT). Esto es sorprendente ya que, a nivel de especie, *Paullinia* y *Citrus* están más estrechamente relacionados entre sí que con *Camellia* o *Coffea*. La evidencia anterior sugiere que tanto las enzimas CS y XMT pueden haber evolucionado de manera convergente para producir cafeína en estas especies (Huang, 2017) (Fig. 12).

2.6.3. MECANISMOS DE DUPLICACIÓN Y FUSIÓN GÉNICAS

Estas estrategias fueron desarrolladas por células ancestrales y evolucionadas en muchas especies a modo de mantener funciones vitales a partir de vías metabólicas que necesitaban de enzimas clave para su supervivencia. Esto puede ser debido a que dichas enzimas ancestrales poseían una amplia preferencia de sustratos que permitían catalizar varias reacciones químicas distintas, por ende, en la evolución de las vías metabólicas se hipotetiza que éstas estaban formadas por una cantidad limitada de enzimas participando de varios procesos biológicos, pero interconectados. Esto dio paso a la expansión de los genomas ancestrales y de las rutas metabólicas mediante duplicaciones masivas y aumento de la especificidad de enzimas por mutaciones puntuales y adaptación (Fani, 2012).

La importancia de la duplicación de genes para el desarrollo de innovaciones metabólicas fue discutida por primera vez por Lewis (1951) y luego por Ohno (1970) y ha sido confirmada recientemente por el análisis comparativo de secuencias completas de genomas de arqueas,

bacteria y eucaria (Fig. 13).

Los genes que descienden de un ancestro común a través de un evento de duplicación se denominan parálogos, y pueden sufrir duplicaciones sucesivas que conducen a una familia de genes parálogos, catalizando reacciones diferentes, aunque con determinadas similitudes.

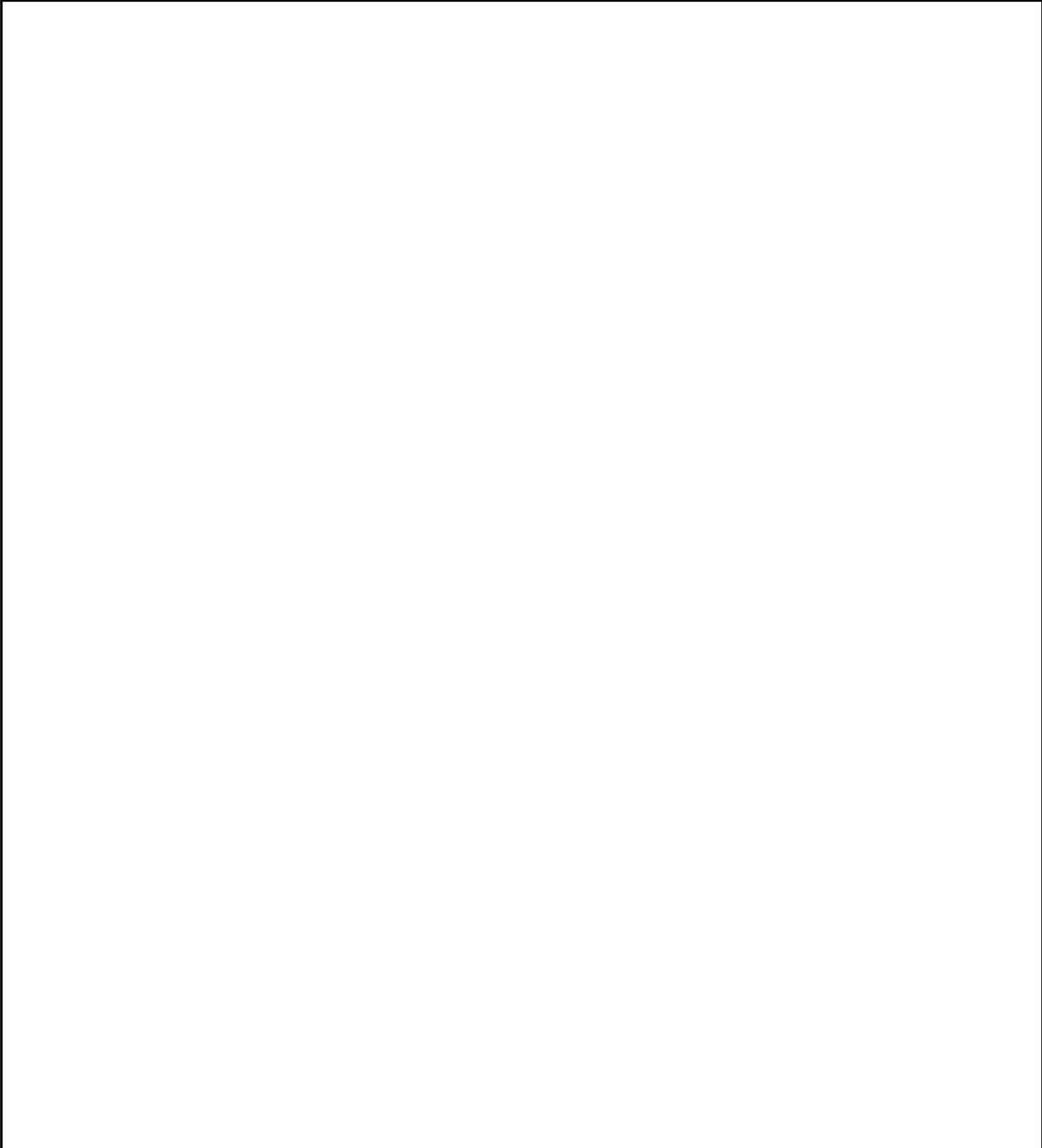


FIG. 12: Árbol filogenético de órdenes de angiospermas correspondientes a Astéridas y Rósidas. Las líneas de colores muestran que al menos 5 veces ha evolucionado la vía metabólica de la cafeína en determinados grupos de plantas de eudicotas.

Los eventos de duplicación pueden generar genes dispuestos en tándem o dispersos en diferentes loci dentro del genoma (Fani 2004; Li y Graur 1991). Existen al menos dos destinos diferentes cuando una duplicación en tándem es generada: 1) los dos genes experimentan una divergencia evolutiva, convirtiéndose en parálogos, y 2) los dos genes se fusionan,

duplicando su tamaño original formando un gen de mayor longitud. Cualquiera de estos dos destinos puede dar lugar a un escenario en mosaico en el cual los genes podrían adquirir funciones novedosas dando lugar a funciones metabólicas distintas.

Las duplicaciones en tándem de ADN a menudo son el resultado de un “*crossing-over*” desigual entre dos moléculas de ADN, pero se pueden invocar otros procesos, como el deslizamiento de la replicación, para explicar la existencia de genes parálogos dispuestos en tándem.

La presencia de genes parálogos en diferentes sitios dentro de un genoma microbiano podría ser el resultado de la actividad antigua de elementos transponibles y/o la duplicación de fragmentos de genoma, así como de duplicaciones del genoma completo (Fani, 2004).

Además de la duplicación génica, otra ruta de evolución génica es la fusión de cistrones independientes que conducen a proteínas bi o multifuncionales (Brilli y Fani 2004; Xie *et al.* 2003). Las fusiones que se han descubierto en genes de muchas rutas metabólicas proporcionan un mecanismo para la asociación física de diferentes dominios catalíticos o de estructuras catalíticas y reguladoras (Jensen, 1976). Las fusiones frecuentemente involucran genes que codifican proteínas que funcionan de manera concertada, como las enzimas que catalizan pasos secuenciales dentro de una vía metabólica (Yanai *et al.* 2002).

La fusión de tales centros catalíticos probablemente promueve la canalización de intermedios que pueden ser inestables o de baja concentración. La alta aptitud de las fusiones de genes también puede depender de la estricta regulación de la expresión de los dominios fusionados. Aunque los eventos de fusión génica se han descrito en muchos procariontes, pueden tener un significado especial entre las células nucleadas, donde el número muy limitado, si no la ausencia completa de operones, no permite la síntesis coordinada de proteínas por ARNm policistrónicos (Fig. 14).

Tesis Doctoral
Sebastián Maximiliano Rossi

FIG. 13: Representación esquemática de la formación de genes parálogos (A) y de familias de genes parálogos (B). **Fuente:** Modificado de Fani, 2004.

FIG. 14: Elongación génica. La duplicación de un gen ancestral y la posterior fusión de los dos homólogos para producir una proteína más larga. **Fuente:** modificado de Fani, 2004.

2.7. VÍA METABÓLICA DE LA CAFEÍNA

2.7.1. LAS XANTINAS

Las xantinas son sustancias que pertenecen a un grupo químico de bases purínicas. Desde el punto de vista médico y farmacológico existen tres xantinas de importancia: la cafeína, la teobromina y la teofilina que son las tres xantinas metiladas, por lo que también se les conoce como metilxantinas (Spiller, 1998; Katzung, 1995). Son consideradas alcaloides, ya que son sustancias fisiológicamente activas, contienen nitrógeno y se encuentran en plantas superiores. Los derivados de la xantina son un grupo de alcaloides comúnmente utilizados por sus efectos como estimulantes leves y broncodilatadores, especialmente en el tratamiento del

asma o los síntomas de la influenza. A diferencia de otros estimulantes más potentes como las aminas simpaticomiméticas, las xantinas actúan principalmente para oponerse a las acciones de la adenosina y aumentar el estado de alerta en el sistema nervioso central (PubChem, National Library of Medicine, US National Institutes of Health. 2019).

2.7.2. METILXANTINAS

Las metilxantinas y los ácidos metilúricos son metabolitos secundarios de plantas que derivan de nucleótidos de purinas (Ashihara y Crozier 1999). Las metilxantinas mejor caracterizadas son la cafeína (1,3,7-trimetilxantina) y la teobromina (3,7-dimetilxantina) que aparecen en té, café, cacao, guaraná, citrus, yerba mate, y en una cantidad de bebidas no alcohólicas de origen en plantas. La cafeína fue aislada primeramente en té y café a comienzos del 1820, pero la principal vía biosintética y catabólica de la cafeína no fue sino establecida recientemente, cuando se pudo obtener cafeína altamente purificada de hojas de té y pudieron clonar los genes que codifican la enzima (Kato *et al.* 1999; Kato *et al.* 2000).

2.7.2.1. DISTRIBUCIÓN

Las metilxantinas han sido encontradas en unas 100 especies distribuidas en 13 órdenes del reino de las plantas (Ashihara y Suzuki 2004; Ashihara y Crozier 1999). Comparado con otros alcaloides de las plantas como la nicotina, morfina y estricnina, los alcaloides de purina están distribuidos ampliamente en el reino de las plantas, aunque la acumulación de altas concentraciones se restringe a un número limitado de especies que incluyen *Coffea arabica* (café), *Camellia sinensis* (té), *Theobroma cacao* (cacao), *Paullinia cupana* (guaraná) e *Ilex paraguariensis*. Todas las plantas que contienen cafeína, excepto la *Scilla marítima*, pertenecen a las Dicotiledoneae. En algunas especies la principal metilxantina es teobromina o ácidos metilúricos, incluyendo teacrina (1,3,7,9-ácido tetrametilúrico) (Fig. 15) en lugar de cafeína (Ashihara y Crozier 1999).

2.7.3. ENZIMAS DE LA FAMILIA SABATH (SALICYLIC ACID, BENZOIC ACID, THEOBROMINE METHYLTRANSFERASE)

Esta vía mediada por enzimas SABATH parece haber evolucionado al menos 5 veces en distintos órdenes durante la historia evolutiva de las angiospermas.

7-Metilxantina Teobromina Cafeína

Teofilina Paraxantina Teacrina

1,3,7-Ácido trimetilúrico Metiliberina Liberina

Metilxantina

3-Metilxantina 1-

FIG. 15: Estructuras de alcaloides de purina presente en plantas que constituyen las metilxantinas

La distribución filogenética de la cafeína o alcaloides de xantina es esporádico y restringido solo a pocos géneros de plantas, como se estableció anteriormente. La cafeína se acumula en varios tejidos con varios fines, los cuales detienen a varios herbívoros o potencian la memoria de los polinizadores. La biosíntesis de cafeína evolucionó independientemente en *Coffea* y *Camellia* involucrando 3 reacciones de metilación. En *Coffea*, por ejemplo, son utilizadas tres enzimas de tipo XMT (metiltransferasas de xantina) de la familia SABATH para catalizar los pasos de metilación de la vía. En *Camellia* se utiliza una vía paráloga, evolucionada convergentemente, enzimas de tipo CS cafeína sintasa.

Numerosos estudios de más de 30 años indican que existen varias rutas de biosíntesis. Esta ruta de biosíntesis de cafeína ha evolucionado independientemente en *Coffea* y en *Camellia* como ha sido detallada anteriormente, involucrando tres reacciones de metilación para convertir secuencialmente xantosina a 7-metilxantosina (1) a teobromina (2) a cafeína (3) (Huang *et. al*, 2016).

2.7.3.1. CARACTERÍSTICAS Y DISTRIBUCIÓN

La mayoría de las enzimas SABATH catalizan la metilación de átomos de oxígeno de una amplia diversidad de ácidos carboxílicos como, por ejemplo, ácido antranílico, benzoico, giberílico, jasmónico, logánico, salicílico y ácido indol-3-acético para fenómenos de floración, defensa, y modulación hormonal. La metilación de átomos de nitrógeno de los alcaloides de xantina mediada por XMT y CS, es probablemente una actividad evolucionada recientemente (Huang *et al*. 2016).

Esta familia multigénica existe en muchas especies de plantas (monocotiledóneas, dicotiledóneas y musgos), y el número de miembros de la familia en estas especies de plantas varía debido a diferentes eventos de duplicación y recombinación de genes. Por ejemplo, *Arabidopsis thaliana* tiene 24 miembros SABATH, en el arroz (*Oryza sativa*) 41 miembros, 33 en el álamo (*Populus trichocarpa*) y solo 4 miembros SABATH en el musgo (*Physcomitrella patens*). Muchos miembros de esta familia se han caracterizado funcionalmente y participan en diversas vías de metabolismo secundario en las plantas, produciendo diferentes tipos de compuestos. Se caracterizaron tres miembros SABATH que tienen estructuras cristalinas proteicas disponibles. Estas estructuras cristalinas son útiles para el modelado de homología

para identificar residuos de aminoácidos importantes y para investigar las interacciones enzimáticas con diferentes sustratos.

La familia SABATH codifica un grupo de metiltransferasas (MT) que pueden transferir un grupo metilo (-CH₃) del donante de metilo, S-adenosil-L-metionina (SAM), ya sea al grupo carboxilo, al grupo azufre, o al anillo de nitrógeno de un determinado sustrato, formando S-adenosil-L-homocisteína (SAH) y éster *O*-metilado, como benzoato de metilo y salicilato de metilo, éster S-metilado, como tiolbenzoato de metilo, o compuestos *N*-metilados como la cafeína.

Los miembros de esta familia que son capaces de metilar el grupo carboxilo de un sustrato dado incluyen al ácido logánico metiltransferasa (LAMT), ácido salicílico metiltransferasa (SAMT), ácido benzoico metiltransferasa (BAMT), ácido jasmónico metiltransferasa (JMT), ácido indol-3-acético metiltransferasa (IAMT), ácido farnesoico metiltransferasa (FAMT), ácido giberélico metiltransferasa (GAMT), cinamato y p-cumarato metiltransferasa (CCMT), ácido antranílico metiltransferasa (AAMT), ácido nicotínico metiltransferasa (NAMT) y ácido p-metoxibenzoico metiltransferasa (MBMT).

El miembro SABATH que metila el grupo azufre de un sustrato dado es la tiol metiltransferasa del musgo recientemente caracterizado. Miembros de SABATH que metilan el anillo de nitrógeno de un sustrato dado incluyen enzimas que participan en la biosíntesis de cafeína, como, por ejemplo, la cafeína sintasa (TCS1) caracterizada en *Camellia*, teobromina sintasa (BTS1) caracterizada en *Theobroma*, 7-metilxantina metiltransferasa (MXMT), xantosina metiltransferasa (XMT) y 3,7 dimetilxantina metiltransferasa (DXMT) encontrada en *Coffea*, así como la enzima es responsable de la biosíntesis de trigonelina.

Los compuestos producidos por los miembros de la familia SABATH juegan papeles importantes en el metabolismo secundario de las plantas que las producen. Por ejemplo, las giberelinas metiladas y el indol-3-acetato regulan el crecimiento de las plantas. El jasmonato de metilo, el salicilato de metilo y la cafeína son producidas en respuestas de defensa de la planta inducidas contra el ataque de patógenos, así como también durante la alimentación de insectos y la formación de algún tipo de daño tisular. El benzoato de metilo, el salicilato de metilo y la cafeína también atraen polinizadores o animales dispersores de semillas.

Los genes del linaje parálogo XMT y CS en la familia SABATH han evolucionado independientemente de manera convergente, porque los genes en los dos clados están involucrados en la biosíntesis de cafeína.

2.8. GENERACIÓN DE LA VÍA DE BIOSÍNTESIS DE CAFEÍNA EN PLANTAS SUPERIORES

A pesar de la existencia de una metiltransferasa SABATH en el musgo *Physcomitrella patens* que cataliza la S-metilación de tioles y tiene un papel en la desintoxicación (Zhao *et al.* 2012) hasta el momento solamente se habían identificado en plantas con semillas y actualmente se ha estudiado de manera exhaustiva en al menos cinco órdenes de plantas superiores tanto en rósidas como en astéridas (Sapindales, *Paullinia* y *Citrus*; Malvales, *Theobroma*, *Cola* y *tilia*; Ericales, *Camellia*; Aquifoliales, *Ilex*; y Gentianales, *Coffea*, cómo las plantas producen cafeína solo se había investigado en detalles en *Camellia* y *Coffea*.

Se ha estudiado esta vía en especies de plantas denotando que las mismas han evolucionado al menos 5 veces durante la historia evolutiva de las angiospermas. La biosíntesis de cafeína evolucionó independientemente en *Coffea* y *Camellia* involucrando 3 reacciones de metilación. En *Coffea*, 3 enzimas de tipo XMT, metiltransferasas de xantina de la familia SABATH son utilizadas para catalizar los pasos de la vía. En *Camellia* se utiliza una vía paróloga, evolucionada convergentemente, enzimas de tipo CS cafeína sintasa (Huang *et al.* 2016). Existen al menos 12 vías alternativas para la producción de cafeína a partir de los sustratos de xantina o xantosina (fig. 16).

2.8.1. FORMACIÓN DE 7-METILXANTINA

La formación de monometilxantina en la ruta biosintética de cafeína principal, se inicia mediante la conversión de xantosina en 7-metilxanthosina. Esta reacción es catalizada por la 7-metilxanthosina sintasa (EC 2.1.1.158).

Los genes que codifican la 7-metilxantosina sintasa, CmXRS1 (AB034699) y CaXMT

(AB048793), se aislaron de *Coffea arabica* (Mizuno *et al.* 2003; Uefuji *et al.* 2003). El segundo paso implica una nucleosidasa que cataliza la hidrólisis de 7-metilxantosina. Se pensó que la N-metilnucleosidasa (EC 3.2.2.25), en las hojas de té, participa en esta reacción (Negishi *et al.*

1988), pero los estudios estructurales sobre la 7-metilxantosina sintasa del café sugirieron que la transferencia de metilo y la escisión de nucleósidos pueden estar acoplados y catalizados por una sola enzima (McCarthy y McCarthy 2007).

FIG. 16: La vía biosintética de la cafeína en las especies de plantas estudiadas que son capaces de producirla, tienen al menos 12 potenciales vías de producción. La única vía bien caracterizada de plantas se muestra con flechas negras e involucra una metilación secuencial de xantosina en el N-7 del anillo heterocíclico como así también a nivel del N-3 de la 7-metilxantina y N-1 de la teobromina. Cada paso de metilación es llevado a cabo por distintas metiltransferasas de alcaloides de xantina en *Coffea*. En contraparte, *Camellia* emplea una enzima relacionada distantemente, la cafeína sintasa, TCS1 para el segundo y tercer paso de metilación, mientras que la enzima que cataliza el primer paso no se encuentra caracterizada. Otras vías bioquímicas potenciales de producción de cafeína se muestran en flechas discontinuas, pero las enzimas especializadas para dichas conversiones son desconocidas. El clivaje de ribosa por 7-metilxantosina no se muestra, pero debería ocurrir de forma concomitante con la metilación de N-7 de xantosina. CF, cafeína; PX, paraxantina; TB, teobromina; TP, teofilina; X, xantina; 1X, 1-metilxantina; 3X, 3-metilxantina; 7X, 7-metilxantina; XR, xantosina (modificado de Huang *et al.* 2016).

2.8.2. FORMACIÓN DE TEOBROMINA

El tercer paso en las vías de biosíntesis de cafeína también está catalizado por la N

metiltransferasa dependiente de S-adenosil-L-metionina (SAM). La cafeína sintasa altamente purificada (EC 2.1.1.160) obtenida de hojas de té jóvenes tiene una amplia especificidad de

sustrato y cataliza la conversión en dos etapas de 7-metilxantina a cafeína a través de teobromina (Kato *et al.* 1999). Esta enzima es distinta de la N-metiltransferasa que cataliza el primer paso de metilación en la vía de la cafeína. El ADNc aislado de las hojas jóvenes de té, denominado TCS1 (AB031280), consta de 1.438 pares de bases y codifica una proteína de 369 aminoácidos (Kato *et al.* 2000). La función de TCS2 (AB031281), que es considerado como un gen parálogo a TCS1 en el genoma del té, aún no se había determinado (Yoneyama *et al.* 2006). Muchos genes que codifican N-metiltransferasas que tienen diferentes especificidades de sustratos, se han aislado de plantas de café.

CCS1 (AB086414), CtCS7 (AB086415) y CaDXMT1 (AB084125) son genes de cafeína sintasa (Mizuno *et al.* 2003; Uefuji *et al.* 2003). Las cafeínas sintasas recombinantes (EC 2.1.1.160) pueden utilizar paraxantina, teobromina y 7-metilxantina como sustratos.

CTS1 (AB034700), CTS2 (AB054841), CaMXMT1 (AB048794) y CaMXMT2 (AB084126) se identificaron como genes que codifican la teobromina sintasa (Mizuno *et al.* 2001; Ogawa *et al.* 2001). La actividad de la teobromina sintasa recombinante (EC 2.1.1.159) es específica para la conversión de 7-metilxantina en teobromina.

La teobromina sintasa, pero no la cafeína sintasa de doble función, parece participar principalmente en la síntesis de teobromina en plantas que acumulan teobromina, como *Theobroma cacao*, *Camellia ptilophylla* y *Camellia irrawadiensis* (Yoneyama *et al.* 2006).

2.8.3. CONVERSIÓN DE TEOBROMINA EN CAFEÍNA

La conversión de teobromina en cafeína es realizada por la cafeína sintasa de doble funcionalidad discutida anteriormente. La metilación del N1 de 7-metilxantina por la cafeína sintasa es mucho más lenta que la del N3 y, como consecuencia, la teobromina se acumula temporalmente en los tejidos sintetizadores de cafeína. Este es el paso final en la ruta principal de biosíntesis de cafeína, es decir, xantosina a 7-metilxantosina a 7-metilxantina a teobromina a cafeína.

Hasta la fecha, se han identificado tres genes de cafeína sintasa en el café (Mizuno *et al.* 2003; Uefuji *et al.* 2003). Los perfiles de expresión de estos genes en diferentes órganos son variables y las propiedades cinéticas de cada enzima recombinante, como los valores de k_m ,

son diferentes. Por lo tanto, las enzimas que participan en la biosíntesis de cafeína en los órganos y en diferentes etapas de crecimiento pueden variar.

Además de la ruta principal de biosíntesis de cafeína, también pueden operar varias rutas menores que dependen principalmente de las amplias especificidades de las N metiltransferasas, especialmente la cafeína sintasa. Como, por ejemplo, la cafeína sintasa cataliza la síntesis de 3-metilxantina a partir de xantina. La paraxantina se sintetiza a partir de 7-metilxantina. Sin embargo, ocurre muy poca acumulación de estos compuestos en tejidos vegetales. La 3-metilxantina puede catabolizarse a xantina y la paraxantina parece convertirse inmediatamente en cafeína. La paraxantina es el sustrato más activo de la cafeína sintasa, pero solo se acumulan cantidades limitadas de paraxantina en los tejidos de las plantas, porque la metilación del N1 de la 7-metilxantina es muy lenta (Ashihara *et al.* 2008).

2.8.4. SUMINISTRO DE XANTOSINA PARA LA BIOSÍNTESIS DE CAFEÍNA

La xantosina, el sustrato inicial de la síntesis de alcaloides de purina, es suministrada por al menos cuatro vías diferentes: biosíntesis de purina de novo (ruta de novo), degradación de nucleótidos de adenina (ruta de AMP), ciclo SAM (ruta de SAM) y nucleótidos de guanina (ruta GMP).

2.8.4.1. RUTA DE NOVO

Al igual que los mamíferos, las plantas sintetizan nucleótidos de purina por vías de rescate y *de novo* (Ashihara y Crozier 1999; Moffatt y Ashihara 2002; Stasolla *et al.* 2003), aunque algunas secciones de las vías son exclusivas de las plantas. La utilización de IMP, formada por la vía biosintética de purina *de novo*, para la biosíntesis de cafeína fue demostrado mediante la utilización de hojas jóvenes de té usando precursores e inhibidores marcados de ^{14}C y ^{15}N glicina de la biosíntesis de purina *de novo* (Ito y Ashihara 1999).

La xantosina es producida por la vía IMP----XMP----Xantosina. La IMP deshidrogenasa (EC 1.1.1.205) y la 5'-nucleotidasa (EC 3.1.3.5) catalizan estas reacciones. La ribavirina, un inhibidor de la deshidrogenasa IMP, reduce la tasa de biosíntesis de cafeína en las plantas de té y café (Keya *et al.* 2003) (Fig. 17).

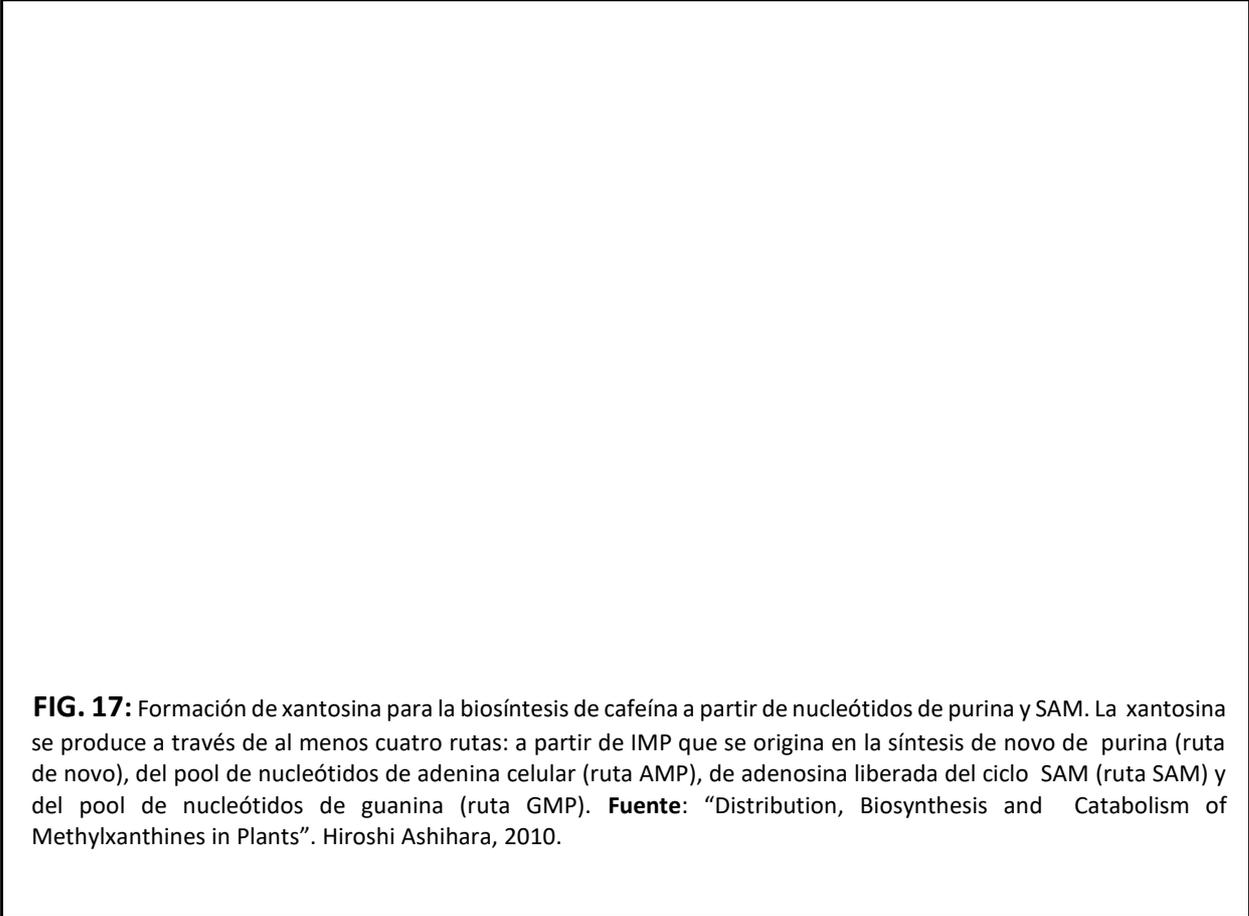


FIG. 17: Formación de xantosina para la biosíntesis de cafeína a partir de nucleótidos de purina y SAM. La xantosina se produce a través de al menos cuatro rutas: a partir de IMP que se origina en la síntesis de novo de purina (ruta de novo), del pool de nucleótidos de adenina celular (ruta AMP), de adenosina liberada del ciclo SAM (ruta SAM) y del pool de nucleótidos de guanina (ruta GMP). **Fuente:** "Distribution, Biosynthesis and Catabolism of Methylxanthines in Plants". Hiroshi Ashihara, 2010.

2.8.4.2. RUTA AMP

Una parte de la xantosina utilizada para la biosíntesis de cafeína se deriva de los conjuntos de nucleótidos de adenina y guanina que se producen mediante las vías *de novo* y de rescate. Hay varias vías potenciales para la síntesis de xantosina a partir de AMP, aunque el AMP—IMP—XMP—ruta de Xantosina, es probable que predomine. Las tres enzimas involucradas en la conversión se han detectado en las hojas de té (Koshiishi *et al.* 2001).

2.8.4.3. RUTA SAM

La ruta SAM es una variación de la ruta AMP. SAM es el donante de metilo para varias reacciones de metilación en la vía biosintética de la cafeína. En el proceso, SAM se convierte

en S-adenosil-L-homocisteína (SAH), que luego se hidroliza en homocisteína y adenosina. La homocisteína se recicla a través del ciclo SAM para reponer los niveles de SAM, y la adenosina liberada del ciclo se convierte en AMP y se utiliza para la biosíntesis de cafeína por la ruta AMP. Dado que se producen 3 moles de SAH a través del ciclo SAM por cada mol de cafeína

que se sintetiza, en teoría esta vía tiene la capacidad de ser la única fuente tanto del esqueleto de purina como de los grupos metilo necesarios para la biosíntesis de cafeína en las hojas jóvenes de té (Koshiishi *et al.* 2001).

2.8.4.4. RUTA GMP

La xantosina utilizada para la biosíntesis de cafeína también se produce a partir de nucleótidos de guanina mediante una vía GMP--guanosina--xantosina. La 5'-nucleotidasa (EC 3.1.3.5) y la guanosina desaminasa (EC 3.5.4.15) participan en esta conversión (Negishi *et al.*, 1994).

2.8.5. N-METILTRANSFERASAS INVOLUCRADAS EN LA BIOSÍNTESIS DE METILXANTINA

2.8.5.1. EXPRESIÓN GÉNICA EN PLANTAS DE CAFÉ Y TÉ

Se ha demostrado la expresión de genes implicados en la biosíntesis de cafeína en hojas jóvenes, botones florales y endospermo en desarrollo de *Coffea arabica* (Mizuno *et al.* 2003). En estudios realizados, se demostró la expresión de CmXRS1, CTS2 y CCS1, que codifican 7-metilxantosina sintasa, teobromina sintasa y cafeína sintasa, respectivamente. En dichos estudios efectuados, los transcritos de CmXRS1 y CCS1 se observaron en todos los órganos, pero el nivel más alto se encontró en el desarrollo del endospermo. También se destacó la presencia de una expresión significativa de CTS2 solo en los botones florales. Los patrones de expresión de CmXRS1 y CCS1 fueron sincronizados. Durante el desarrollo de los frutos de *Coffea arabica*, los transcritos de CmXRS1 y CCS1 estaban presentes en todas las etapas de crecimiento, excepto en tejidos completamente maduros. El patrón de expresión de estos genes durante el crecimiento está más o menos relacionado con la síntesis *in situ* de cafeína a partir de nucleótidos de adenina, aunque se encontraron excepciones en las etapas muy tempranas y tardías del crecimiento del fruto. Dado que el nivel de transcritos CTS2 que

codifican la teobromina sintasa es muy bajo en las frutas, el gen CCS1 alternativo que codifica la cafeína sintasa de doble función puede ser operativo para los dos últimos pasos de la biosíntesis de cafeína. En el desarrollo de frutos de *Coffea arabica*, los niveles de transcritos de CmXRS1 y CCS1 son más altos en semillas que en pericarpio. La actividad nativa de la cafeína sintasa (3N-metiltransferasa) se distribuye en ambos órganos de manera similar. Por lo tanto,

la cafeína que se acumula en las semillas de café maduras parece sintetizarse dentro de las semillas en desarrollo y no se transporta desde el pericarpio (Koshiro *et al.* 2006).

En *Camellia sinensis*, la expresión de cafeína sintasa que codifica TCS1 es mayor en las hojas jóvenes que en las hojas, tallos o raíces maduras (Li *et al.* 2008). Esto es consistente con el hecho de que la actividad biosintética de la cafeína ocurre principalmente en las hojas jóvenes (Ashihara y Kubota 1986). Estudios recientes que utilizan el cultivo de tejidos de *Camellia sinensis* indican que la expresión de TCS1, y posiblemente el gen no identificado que codifica la 7-metilxantosina sintasa, representa el mecanismo de control principal para la biosíntesis de cafeína. Aunque se observó un mayor contenido de cafeína cuando se pusieron a crecer cultivos en medios que contenían paraxantina, la adición de adenosina, guanosina o hipoxantina no tuvo un impacto similar. Por lo tanto, ni el suministro de precursores de purina no metilados ni la disponibilidad de SAM parecen ser un factor importante en la regulación de la biosíntesis de cafeína (Li *et al.* 2008; Deng *et al.* 2008).

2.8.5.2. RELACIÓN EVOLUTIVA DE LA CAFEÍNA SINTASA Y ENZIMAS RELACIONADAS

Hay cuatro regiones altamente conservadas: motivo A, motivo B', motivo C y la región YFFF en la secuencia de aminoácidos de la familia de cafeína sintasa (Kato y Mizuno, 2004) (Fig. 18). Tres motivos conservados, A, B y C, del sitio de unión del donante de metilo de SAM se ha informado en la mayoría de las plantas que son O-metiltransferasas dependientes de SAM (Joshi y Chiang, 1998).

El motivo B' y la región YFFF contienen muchos aminoácidos hidrofóbicos que son específicos

del motivo B' de la familia metiltransferasa. La mayoría de los miembros de esta familia metiltransferasa con motivos B' caracterizados, catalizan la formación de pequeños y volátiles ésteres metílicos utilizando SAM como donador de metilo y sustratos con un grupo carboxilo como aceptor de metilo. Los miembros de esta familia incluyen ácido salicílico carboxil metiltransferasa (SAMT) (Ross *et al.* 1999), ácido benzoico carboxil metiltransferasa (BAMT) (Dudareva *et al.* 2000), ácido jasmónico carboxil metiltransferasa (JAMT) (Seo *et al.* 2001), ácido farnesoico carboxil metiltransferasa (FAMT) (Yang *et al.* 2006), ácido indol-3-acético metiltransferasa (IAMT) (Zhao *et al.* 2008), ácido giberélico metiltransferasa (GAMT)