

AUTENTICACIÓN DEL ORIGEN BOTÁNICO Y GEOGRÁFICO DEL CAFÉ COLOMBIANO POR METABOLÓMICA BASADA EN RMN

VICTORIA ANDREA ARANA RENGIFO



**UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS
PROGRAMA DE DOCTORADO EN CIENCIAS - QUÍMICA**

**SANTIAGO DE CALI
2015**

AUTENTICACIÓN DEL ORIGEN BOTÁNICO Y GEOGRÁFICO DEL CAFÉ COLOMBIANO POR METABOLÓMICA BASADA EN RMN

VICTORIA ANDREA ARANA RENGIFO

**Proyecto de Trabajo de Grado presentado como requisito para optar al título de
Doctora en Ciencias - Química**

**Director
JULIEN WIST, Ph.D.**

**UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS
PROGRAMA DE DOCTORADO EN CIENCIAS - QUÍMICA**

**SANTIAGO DE CALI
2015**

UNIVERSIDAD DEL VALLE
FACULTAD DE CIENCIAS NATURALES Y EXACTAS

**AUTENTICACIÓN DEL ORIGEN BOTÁNICO Y GEOGRÁFICO DEL CAFÉ
COLOMBIANO POR METABOLÓMICA BASADA EN RMN**

VICTORIA ANDREA ARANA RENGIFO

Palabras claves: Café, Arábica, Robusta, huella metabolómica, ^1H -NMR, GC-MS, GC-C-IRMS, quimiometría, origen geográfico, origen botánico.

DEDICATORIA

A mis padres, Blanca Victoria Rengifo y Marino Orlando Arana, por su apoyo constante, comprensión e infinito amor.

A mi amado esposo e hijo, Jorge Enrique Trilleras y Camilo Trilleras, por hacer parte de mis sueños y bendiciones.

AGRADECIMIENTOS

Me gustaría expresar mis agradecimientos a todas las personas que de alguna manera han contribuido a la realización de esta Tesis.

A mi familia, mis padres y mi esposo, por siempre creer en mi y ayudarme a recorrer este camino con su apoyo constante.

A mi director de Tesis, Julien Wist, por la confianza que depositó en mi y en el desarrollo de esta investigación, y la oportunidad que me brindo al pertenecer al grupo de investigación en Desarrollo y Aplicaciones en Resonancia Magnética Nuclear (DARMN) en donde he crecido tanto profesional como personalmente.

A Edgar Moreno y Rodrigo Alarcon, de ALMACAFE, por todo el apoyo y colaboración con las muestras de café que necesitamos para la ejecución de esta investigación.

A la Dra Helena Maruenda Castillo, Dr. Luis Norberto Benítez Vásquez y Dr. Daniel Ricardo Molina Velasco por su dedicación en la evaluación de esta Tesis.

A mis amigas, Jessica Medina y Ana Karina Velez por su maravillosa amistad y estar siempre presente en todos los momentos. A mis amigos, Marlon Doney Martinez, Antonio José Garzón, Cristian Pantoja y Alejandro Bolaños por toda su ayuda.

A Sara Pereddo, Edison Castro, Angela María Ormaza y Andrea Molina por todos los momentos compartidos.

A Liberney Osorio por ayudarme en todo este proceso, desde el inicio hasta el final.

A Pierre Esseiva, Juan Manuel García y Diego Pazos, del Instituto de Policía Científica, Universidad de Lausanne - Suiza, por brindarme toda su colaboración para realizar los análisis CG-MS y GC-C-IRMS en las muestras de café durante mi estancia doctoral.

A Carlos Martín, Julia Colmenares, Laura Victoria Suárez, Juan David Villegas, Diego Merchan, Andres Castillo y Andres Bernal, por hacer de mi estancia en Suiza un recuerdo inolvidable.

A Luis Eduardo Hurtado y Carlos Rodriguez, del Laboratorio de RMN, por darme el espacio en el equipo de RMN y brindarme su experiencia y amistad.

A la Dra Irma Janeth Sanabría y Dany Mercedes Acevedo, del Grupo de Investigación de Procesos Avanzados para Tratamientos Físicos y Químicos (GAOX), por facilitarme las instalaciones y equipos de su laboratorio para realizar el tratamiento de las muestras de café.

TABLA DE CONTENIDO

LISTA DE FIGURAS	ix
LISTA DE TABLAS	xii
LISTA DE ANEXOS	xiii
LISTA DE SÍMBOLOS, ABREVIATURAS Y/O ACRÓNIMOS	xiv
RESUMEN	xvi
ABSTRACT	xviii

1. INTRODUCCIÓN	1
------------------------	----------

1.1	CULTIVO DE CAFÉ	1
1.2	COMPOSICIÓN QUÍMICA DEL CAFÉ VERDE	2
1.3	CAMBIOS EN LA COMPOSICIÓN QUÍMICA DEL CAFÉ VERDE DURANTE EL PROCESO DE TOSTADO	8
1.4	CAFÉ DE COLOMBIA	11
1.5	PLATAFORMAS ANALÍTICAS	13
1.6	METABOLÓMICA BASADA EN RMN	16
1.7	PROCESAMIENTO DE LOS DATOS	21
1.8	ANÁLISIS ESTADÍSTICO MULTIVARIANTE	21

2. METODOLOGÍA	30
-----------------------	-----------

2.1	RECOLECCIÓN DE MUESTRAS	30
2.1.1	MODELO DE CLASIFICACIÓN DEL ORIGEN BOTÁNICO Y GEOGRÁFICO DE CAFÉ COLOMBIANO	30
2.1.2	EVALUACIÓN DEL EFECTO DEL GRADO DE TOSTADO SOBRE EL MODELO DE CLASIFICACIÓN	31
2.1.3	EVALUACIÓN DE LAS TÉCNICAS ANALÍTICAS GC-MS, GC-C-IRMS Y RMN PARA TERMINAR EL ORIGEN GEOGRÁFICO DEL CAFÉ	32
2.2	DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR RMN	33
2.3	DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR GC-MS	34
2.4	DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE	34

	METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR GC-C-IRMS	
2.5	ANÁLISIS RMN	35
2.6	ANÁLISIS GC-MS	37
2.7	ANÁLISIS GC-C-IRMS	38
2.8	PROCESAMIENTO DE LOS DATOS OBTENIDOS POR RMN	39
2.9	ANÁLISIS MULTIVARIADO	40
2.9.1	ANÁLISIS NO SUPERVISADO	40
2.9.2	ANÁLISIS SUPERVISADOS – MODELOS DE CLASIFICACIÓN	40
2.9.3	ANÁLISIS SUPERVISADOS – MODELOS DE CUANTIFICACIÓN	42
3.	RESULTADOS Y DISCUSIÓN	43
3.1	DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR RMN	43
3.2	HUELLA METABOLÓMICA ¹ H-RMN DE CAFÉ VERDE Y TOSTADO	47
3.3	ANÁLISIS NO SUPERVISADO	54
3.4	ANÁLISIS SUPERVISADO	58
3.4.1	MODELO DE CLASIFICACIÓN	58
3.4.2	MODELO DE CUANTIFICACIÓN	63
3.4.3	VARIABLES RELEVANTES	64
3.5	EVALUACIÓN DEL EFECTO DEL GRADO DE TOSTADO SOBRE LOS MODELOS DE CLASIFICACIÓN	71
3.6	EVALUACIÓN DE LAS TÉCNICAS ANALÍTICAS GC-C-IRMS, GC-MS Y ¹ H-RMN PARA DETERMINAR EL ORIGEN GEOGRÁFICO DEL CAFÉ	77
3.6.1	ANÁLISIS POR GC-C-IRMS	78
3.6.2	ANÁLISIS POR GC-MS	79
3.6.3	ANÁLISIS POR RMN	80
4.	CONCLUSIONES	83
5.	REFERENCIAS	86
6.	ANEXOS	103

LISTA DE FIGURAS

Figura 1.1	Estructura química de la cafeína.	3
Figura 1.2	Estructura química de la trigonelina y ácido nicotínico.	3
Figura 1.3	Ácidos clorogénicos y compuestos relacionados.	5
Figura 1.4	Producción mundial de café en miles de sacos x 60 kg y por año de cosecha (1990 – 2012), para los países de Colombia y Vietnam.	12
Figura 1.5	Secuencia de pulso de Presaturación-Excitación.	18
Figura 1.6	Secuencia de pulso de 1D-NOESY-Presat.	19
Figura 1.7	Secuencia de pulso de 1D-CPMG con presaturación.	20
Figura 1.8	Esquema simplificado de doble validación cruzada.	26
Figura 2.1	Distribución geográfica del origen de las muestras utilizadas para crear un modelo de café Colombiano.	31
Figura 2.2	Grado de tostado del café.	32
Figura 2.3	Región de la supresión de la señal del agua en un extracto de café verde obtenido con solución tampón.	36
Figura 2.4	Región de la supresión de las señales del metanol y agua en un extracto de café tostado.	37
Figura 3.1	Espectros de extractos de café verde obtenidos con los solventes óxido de deuterio, metanol-d ₄ , dimetilsulfóxido-d ₆ y cloroformo-d ₁ .	44
Figura 3.2	Estandarización por ³¹ P-RMN de una nueva solución tampón de fosfato.	46
Figura 3.3	Espectro ¹ H-RMN de una muestra de café verde obtenido con solución tampón y asignación de las principales moléculas.	48
Figura 3.4	Espectro ¹ H-RMN de una muestra de café verde obtenido con metanol y asignación de las principales moléculas.	49
Figura 3.5	Espectro ¹ H-RMN de una muestra de café tostado obtenido con solución tampón acuosa y la asignación de las principales moléculas.	50
Figura 3.6	Espectro ¹ H-RMN de una muestra de café tostado obtenido con metanol y la asignación de las principales moléculas.	51
Figura 3.7	Perfil metabolómico ¹ H-RMN de las especies <i>C. arabica</i> y <i>C. canephora</i> en extractos metanólicos de semillas tostadas.	51
Figura 3.8	Perfil metabolómico ¹ H-RMN del extracto metanólico de café Arábica mezclado con café Robusta en presentación tostado.	52
Figura 3.9	Perfil metabolómico ¹ H-RMN de café arábigo	53

	Colombiano y café arábigo de otros orígenes en extractos metanólicos de semillas tostadas.	
Figura 3.10	Perfil metabolómico ^1H -RMN de la especie <i>C. arabica</i> y <i>C. arabica</i> descafeínada en extractos metanólicos de semillas de café tostado.	54
Figura 3.11	Gráfico de puntuaciones del PCA de los datos espectrales de extractos metanólicos de semillas verdes de <i>C. arabica</i> y <i>C. canephora</i> .	55
Figura 3.12	Gráfico de puntuaciones del PCA de los datos espectrales de extractos metanólicos de semillas tostadas de <i>C. arabica</i> y <i>C. canephora</i> .	56
Figura 3.13	Gráfico de puntuaciones del PCA de los datos espectrales de extractos metanólicos de semillas verdes de <i>C. arabica</i> de origen colombiano y otros orígenes.	57
Figura 3.14	Comparación de los espectros ^1H -RMN de la muestra de café Arábica AC1337_BRAAS10080 que resultó del PCA como dato atípico con una muestra representativa del grupo <i>C. arabica</i> .	58
Figura 3.15	Representación del bucle de doble validación cruzada usado en el modelo estadístico de clasificación de origen botánico y geográfico para café verde y tostado.	59
Figura 3.16	<i>Y predicho vs Y observado</i> del contenido porcentual de <i>C. robusta</i> en muestras de <i>C. arábica</i> tostadas obtenido para los conjuntos de entrenamiento y de prueba usando el modelo PLS.	64
Figura 3.17	Gráfico de puntuaciones y cargas correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen botánico entre “ <i>C. arabica</i> ” y “ <i>No C. arabica</i> ” de café verde.	65
Figura 3.18	Gráfico de puntuaciones y cargas correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen botánico entre “ <i>C. arabica</i> ” y “ <i>No C. arabica</i> ” de café tostado.	66
Figura 3.19	Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen botánico de café verde.	67
Figura 3.20	Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen botánico de café tostado.	67
Figura 3.21	Gráfico de puntuaciones y cargas correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen geográfico entre “ <i>café Colombiano</i> ” y “ <i>otros orígenes</i> ” de café verde.	69
Figura 3.22	Gráfico de puntuaciones y cargas correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen geográfico entre “ <i>café Colombiano</i> ” y “ <i>otros orígenes</i> ” de café tostado.	69

Figura 3.23	Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen geográfico de café verde.	70
Figura 3.24	Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen geográfico de café tostado.	70
Figura 3.25	Superposición de los espectros ^1H -RMN de los extractos metanólicos de una muestra de café con cuatro diferentes grados de tostado procesada a una temperatura de 200 °C.	72
Figura 3.26	Gráfico de puntuaciones del PCA de los datos espectrales de extractos metanólicos de muestras de café tostado con diferentes grados de tostado a 200 °C y 180 °C.	73
Figura 3.27	Gráfico de puntuaciones y cargas obtenidos del modelo PLS-DA para grado de tostado medio vs muy tostado, grado de tostado medio vs tostado oscuro, y grado de tostado medio vs tostado oscuro medio.	74
Figura 3.28	<i>Y predicho vs Y observado</i> de la variable Luminosidad (*L) para predecir el grado de tostado del café obtenido para los conjuntos de entrenamiento y de prueba usando el modelo PLS.	77
Figura 3.29	Diagrama de caja de los datos de la relación $\delta^{13}\text{C}$ de cafeína obtenidos por GC-C-IRMS.	78
Figura 3.30	Gráfico de puntuaciones y cargas del análisis OPLS-DA de muestras de café obtenidos por GC-MS	80
Figura 3.31	Gráfico de puntuaciones del análisis OPLS-DA para la clasificación de café con huella metabolómica ^1H -RMN de acuerdo a los tres países de origen: Colombia, Brasil y Perú.	81
Figura 3.32	Gráfico de puntuaciones y cargas del análisis OPLS-DA de muestras de café obtenidos con huella metabolómica ^1H -RMN.	82

LISTA DE TABLAS

Tabla 1.1	Composición química de semillas verdes de <i>Coffea arabica</i> y <i>Coffea canephora</i> .	2
Tabla 1.2	Composición química de semillas tostadas de <i>Coffea arabica</i> y <i>Coffea canephora</i> .	8
Tabla 1.3	Enfoques metabolómicos para evaluar la autenticidad del café publicados en los últimos 20 años.	15
Tabla 3.1	Valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café verde.	60
Tabla 3.2	Valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café tostado.	62
Tabla 3.3	Valores medios de los parámetros de validación para evaluar el efecto del proceso de tostado en la predicción del origen geográfico.	76

LISTA DE ANEXOS

Anexo 6.1	Muestras de café, verde y tostado, utilizadas en el estudio.	103
Anexo 6.2	Protocolo para la preparación de solución tampón de fosfato y protocolo para la preparación de muestras de café verde y tostado en solución tampón.	106
Anexo 6.3	Protocolo para la preparación de muestras de café verde y tostado en metanol.	109
Anexo 6.4	Protocolo para la preparación de muestras de café tostado para análisis por GC-MS.	111
Anexo 6.5	Protocolo para la preparación de muestras de café tostado para análisis por GC-C-IRMS.	113
Anexo 6.6	Protocolo para la obtención de espectros ^1H -RMN de muestras de café verde y tostado.	115
Anexo 6.7	Protocolo para análisis de extractos de café tostado por GC-MS.	118
Anexo 6.8	Protocolo para análisis de extractos de café tostado por GC-C-IRMS.	119
Anexo 6.9	Predicción de los modelos de clasificación PLS-DA.	120
Anexo 6.10	Datos de los metabolitos obtenidos por GC-C-IRMS y GC-MS en muestras de café tostado para los tres países de estudio.	124
Anexo 6.11	Publicaciones y comunicaciones en congresos	126

LISTA DE SÍMBOLOS, ABREVIATURAS Y/O ACRÓNIMOS

msnm	metros sobre el nivel del mar
FCN	Federación Nacional de Cafeteros de Colombia
IGP	Indicación Geográfica Protegida
DOP	Denominación de Origen Protegida
MC	Marca de Certificación
ALMACAFE	Almacenes Generales de Depósito de Café
FT-MIR	Fourier Transformation Mid-Infrared Reflectance Spectroscopy
DRIFT	Diffuse Reflectance Infrared Fourier Transform Spectroscopy
RBF	Radial-Basis Function
ATR	Attenuated Total Reflection
NIRS	Near-Infrared Spectroscopy
OSC	Orthogonal Signal Correction
DOSC	Direct Orthogonal Signal Correction
OWAVEC	Orthogonal Wavelet Correction
GC	Gases Chromatography
LC	Liquid Chromatography
HPLC	High-Performance Liquid Chromatography
LC-MS	Liquid Chromatography - Mass Spectroscopy
GC-MS	Gases Chromatography - Mass Spectroscopy
GC-C-IRMS	Gases Chromatography - Combustion - Isotope-ratio mass spectrometry
GC-FID	Gases Chromatography - Flame Ionization Detector
ESI FT-ICR MS	Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
HS PTR-MS	High Sensitivity Proton Transfer Reaction Mass Spectrometry
PTR-TOF-MS	Transfer Reactor Time-of-Flight Mass Spectrometry
NMR	Nuclear Magnetic Resonance
RMN	Resonancia Magnética Nuclear
HR-MAS NMR	High Resolution-Magic Angle Spinning Nuclear Magnetic Resonance
IRMS	Isotope-ratio mass spectrometry
PCA	Principal Component Analysis
PCs	Principal Components
LDA	Linear discriminant analysis
ANN	Artificial Neural Network
FDA	Factorial Discriminant Analysis
PLS-DA	Partial Least Squares Discriminant Analysis

OPLS-DA	Orthogonal Partial Least Square Discriminant Analysis
CART	Classification trees
16-OMC	16- <i>O</i> -metilcafestol
1D-NOESY-presat	1D Nuclear Overhauser effect spectroscopy with presaturation
1D-CPMG	1D Carr–Purcell–Meiboom–Gill
HCA	Hierarchical Cluster Analysis
PCR	Principal Components Regression
LDA	Linear Discriminant Analysis
PRESS	Predicted Residual Error Sum of Squares
MSE	Mean Squared Error
TSS	Total Sums of Squares
CV	Cross Validation
<i>n</i> CV	Double Cross Validation
FID	Free Induction Decay
TIC	Total Ion Current
CIE	Commission Internationale de l’Eclairage - Comisión Internacional de Iluminación
ACQ	Ácido cafeoilquínico
AFQ	Ácido feruloilquínico
AdiCQ	Ácido dicafeoilquínico
ACoQ	Ácido <i>p</i> -coumaroilquínico
TMS	Tetrametilsilano

RESUMEN

El Café de Colombia es 100% Arábico lavado (*Coffea arabica* L.). Este, debido a las condiciones ambientales y al sistema de gestión de calidad que han diseñado los productores colombianos, cuenta con una gran reputación a nivel mundial como producto de alta calidad. Se sabe que alrededor de 563.000 familias de cafeteros se han organizado a través de la Federación Nacional de Cafeteros para aplicar normas de calidad para la exportación de café. Para garantizar la autenticidad y origen del Café de Colombia se emplea certificaciones como las de Denominación de Origen, Indicación Geográfica Protegida o Marca de Certificación. Sin embargo, Colombia aún no se vale de plataformas analíticas para demostrar que el producto es realmente colombiano, de calidad y que éste no se encuentra adulterado.

En este trabajo se ha desarrollado un método simple, rápido y robusto, haciendo uso de huella metabolómica basada en ^1H -RMN, para la clasificación del café verde y tostado, por origen botánico (*C. arabica* vs No *C. arabica*) y geográfico (café Colombiano vs cafés de otros orígenes). Los modelos que se obtuvieron por PLS-DA, evaluados mediante doble validación cruzada, permitieron predicciones muy precisas ($Q^2 > 0.74$ para la validación interna y $Q^2 > 0.86$ para la validación externa). Estos modelos para el caso de café tostado fueron construidos y evaluados exclusivamente con muestras preparadas de manera estandarizada, con grado de tostado medio a 200 °C.

Debido a que la composición química del café depende del tiempo y temperatura usada durante el proceso de tostado, en este trabajo también se evalúa el efecto del proceso de tostado sobre la predicción de los modelos. Para ello se usaron, en la validación externa, muestras preparadas con cuatro diferentes grados de tostado,

medio, oscuro medio, oscuro y muy oscuro, y dos temperaturas, 180 y 200 °C. En el caso de las muestras con grado de tostado muy oscuro, se observó que el perfil metabolómico es modificado drásticamente, y los valores de predicción obtenidos fueron muy bajos ($Q^2 < 0.20$). Para los demás grados de tostado, se pueden realizar predicciones cercanas a las obtenidas ($Q^2 > 0.75$), si se elimina de la matriz de datos una parte del área de la región alifática del espectro. De acuerdo a estos resultados, las muestras con grado de tostado muy oscuro fueron detectadas y excluidas del análisis de clasificación, obteniéndose de esta manera un modelo robusto que permite predecir muestras de café tostado con tres grados de tostado diferentes, sin necesidad de modificar el conjunto de muestras usado para la construcción de los modelos.

Colombia exporta, casi en su totalidad, el café que produce por el gran valor con que éste cuenta en el mercado internacional, por consiguiente, necesita importar café de menor calidad para abastecer el consumo interno del país. Esto facilita la ocurrencia de eventos de adulteración del café Colombiano por parte de los caficultores. Por ello, se incluyeron en el estudio muestras de café de países vecinos (Brasil y Perú) y se compararon los resultados de clasificación de estas muestras con respecto a las de Café de Colombia empleando tanto los estudios de metabolómica ^1H -RMN mencionados, como otras técnicas analíticas tales como GC-MS y GC-C-IRMS. Los mejores resultados se obtuvieron con RMN seguido de GC-MS. La metodología GC-C-IRMS no permitió una correcta clasificación.

ABSTRACT

Colombian coffee is 100% washed Arabica (*Coffea arabica* L.) with a worldwide reputation for being a high quality product, it could be said that Colombia's coffee growing regions are characterized by environmental conditions and system of quality management designed by Colombian producers, about 563.000 coffee families have been organized by the National Federation of Coffee Growers to implement quality standards for the export of coffee. Therefore, there are various instruments that provide a guarantee of origin to the Colombian Coffee, as Designation of Origin, Protected Geographical Indication or certification mark in dozens of countries, but for Colombian Coffee it is not enough to be a worldwide reference associated to high standards of quality, it is necessary to demonstrate that a product is really Colombian, in order to ensure quality and detect possible adulterations using analytical platforms.

A simple, fast and robust method was developed for the classification of coffee for its botanical (*C. arabica* vs Not *C. arabica*) and geographical (Colombian coffee vs coffees from other origin) origin in green and roasted coffee, for metabolomics based on ^1H -NMR fingerprinting. The models obtained by PLS-DA and evaluated by double validation cross, allow for very accurate predictions ($Q^2 > 0.74$ for an internal validation set and $Q^2 > 0.86$ for an external validation set) but were exclusively built using samples prepared in a standardized manner, medium roasting at a temperature of 200 °C. The chemical composition of coffee depends on the time and temperature used during the roasting process, thus affecting the overall profile. In order to evaluate the effect of the roasting parameters on the classification, samples were prepared at four different levels of roasting, medium, medium dark, dark and very dark and two different temperatures, 180 and 200 °C. For very dark roasting samples, the prediction accuracy estimated with parameter Q^2 is very poor ($Q^2 < 0.20$), while

for the other degrees of roasting, the predictions are similar to those obtained ($Q^2 > 0.75$) removing of the data matrix one área in the aliphatic región of the spectra. Very dark roasting samples can be detected and thus excluded before proceeding with determining their origin, obtaining a robust model to predict roasted coffee samples with three different degrees of roasting, without the need to modify the set of samples used for contruction models.

Colombia imports large quantities of coffee for its internal consumption, since it obtains high prices for its coffee on the international market. This situation makes imperative the implementation of methods able to detect frauds, i.e., adulterated coffees. Classification results of samples of Colombian coffee against coffees from other neighboring countries (Brazil and Peru) were compared using analytical techniques GC-MS, GC-C-IRMS and $^1\text{H-NMR}$. The results show that the number of variables plays an important role. Indeed, when comparing single techniques, NMR performed best, followed by GC-MS, while no classification could be obtained from GC-C-IRMS.

1. INTRODUCCIÓN

1.1 CULTIVO DE CAFÉ

Café es el nombre común de las plantas de la familia *Rubiaceae*, del género *Coffea*, y comprende más de 90 especies diferentes, solo dos de las cuales, *Coffea arabica* L., (conocida como café Arábica) y *Coffea canephora* P., (conocida como café Robusta) poseen importancia comercial. El café Arábica representa el 75 % de la producción mundial de café, debido a sus cualidades de sabor más pronunciado y fino, es considerado de mejor calidad; y en consecuencia su precio es más alto. Mientras que el café Robusta ofrece una bebida más rica en cafeína; con un sabor más fuerte, amargo y ácido, usualmente es utilizado para la fabricación de café soluble o instantáneo, y mezclas (Tornincasa *et al.*, 2010). Además, las dos especies presentan muchas diferencias en su composición química (Flament, 2002).

Los frutos de café comercialmente son llamados “cerezas”, los granos de café son las semillas de los frutos. Cada cereza contiene en general dos granos de café. Después de la cosecha, los granos de café se extraen rápidamente de la fruta para evitar procesos fermentativos de la pulpa que puede dañar la integridad de la semilla con consecuencias negativas sobre la calidad del café. La eliminación de la pulpa y de los diferentes recubrimientos que rodean a los granos de café se realiza mediante el proceso llamado beneficio, el cuál tiene dos métodos de tratamiento alternativos: el proceso “húmedo”, que produce el café “lavado”, y el proceso “seco”, que produce el café “natural”. El beneficio consiste de una serie de etapas críticas (despulpado, fermentación, lavado y secado) que entran en juego en los atributos finales de la taza. Controlando variables como temperatura, duración y sistema de fermentación

(sumergido en agua o seco) el productor puede elegir los matices de sabor y aroma que tendrá la taza de café (Puertas, 2010; Puertas, 2012). Finalizado el proceso de beneficio, los granos de café sin tostar son llamados café verde. Después el café es transportado a los países consumidores, donde generalmente son mezclados y posteriormente tostados para producir el característico sabor y aroma asociado con la bebida de café.

1.2 COMPOSICIÓN QUÍMICA DEL CAFÉ VERDE

La fracción no volátil de café verde está compuesta principalmente de agua, carbohidratos, proteínas, aminoácidos libres, lípidos, minerales, ácidos orgánicos, ácidos clorogénicos, trigonelina y cafeína (Tabla 1.1).

Tabla 1.1. Composición química de semillas verdes de *Coffea arabica* y *Coffea canephora*.

Componente	Concentración (%)	
	<i>Coffea arabica</i>	<i>Coffea canephora</i>
<i>Carbohidratos</i>		
Sucrosa	6.0 - 9.0	0.9 - 4.0
Azúcares reductores	0.1	0.4
Polisacáridos	34 - 44	48 - 55
<i>Compuestos nitrogenados</i>		
Cafeína	0.9 - 1.3	1.5 - 2.5
Trigonelina	0.6 - 2.0	0.6 - 0.7
Proteínas/péptidos	10.0 - 11.0	11.0 - 15.0
Aminoácidos libres	0.5	0.8 - 1.0
<i>Lípidos</i>		
Aceite de café (triglicéridos con insaponificables, esteroides/tocoferoles)	15.0 - 17.0	7.0 - 10.0
Diterpenos (libres y esterificados)	0.5 - 1.2	0.2 - 0.8
<i>Minerales</i>	3.0 - 4.2	4.4 - 4.5
<i>Ácidos y esteres</i>		
Ácidos clorogénicos	4.1 - 7.9	6.1 - 11.3
Ácidos alifáticos	1.0	1.0
Ácido quínico	0.4	0.4

Fuente: Clarke & Macrae (1985), Clifford (2000), Trugo & Macrae (1984), Trugo (2003), Kölling-Speer & Speer (2005), Speer & Kölling-Speer (2006), Farah, de Paulis, Moreira, Trugo, & Martin

(2006a), Farah & Donangelo (2006b), Holscher, Vitzthum, & Steinhart (1990), and Fischer, Reimann, Trovato, & Redgwell (2001).

Cafeína

La cafeína (Figura 1.1) es una metilxantina con características amargas; sin embargo, es responsable de no más del 10 % del sabor amargo percibido de la bebida de café (Macrae, 1985; Farah, de Paulis, Moreira, Trugo, & Martin, 2006a). Este alcaloide es estable al calor, y su concentración en *C. canephora* es aproximadamente el doble de la que se encuentra en *C. arabica* (Tabla 1.1).

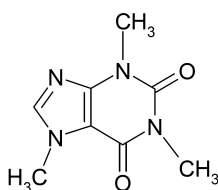


Figura 1.1. Estructura química de la cafeína.

Trigonelina

Trigonelina (Figura 1.2) es un alcaloide biológicamente derivado de la metilación enzimática del ácido nicotínico. Contribuye al sabor de la bebida y es un precursor para la formación de diferentes clases de compuestos volátiles durante el proceso de torrefacción como pirroles y piridinas, algunos de los cuales puede conferir un “sabor desagradable” (Macrae, 1985; Moreira, Trugo, & Maria, 2000). El contenido de trigonelina también se ha correlacionado con la buena calidad en taza (Farah, Monteiro, Calado, Franca, & Trugo, 2006). La cantidad de trigonelina en *C. canephora* es aproximadamente dos terceras partes de la que se encuentra en *C. arabica* (Tabla 1.1).

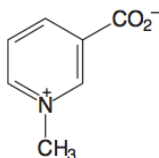


Figura 1.2. Estructura química de la trigonelina.

Ácidos clorogénicos

Los ácidos clorogénicos (Figura 1.3) comprenden una clase importante de compuestos fenólicos, que se derivan principalmente de la esterificación de los ácidos trans-cinámico (por ejemplo, cafeico, ferúlico, y p-cumárico) con (-) - ácido quínico. Se subdividen de acuerdo con la naturaleza y el número de sustituyentes cinámicos, y la posición de esterificación en el anillo de ciclohexano del ácido quínico (Clifford, 2000). Los ésteres se forman preferentemente con el hidroxilo situado en el carbono 5, así como los situados en los carbonos 3 y 4. Las principales subclases de ácidos clorogénicos en el café verde son ácidos cafeoilquínicos (ACQ), ácidos dicafeoilquínicos (AdiCQ), ácidos feruloilquínicos (AFQ) y, con menor abundancia, ácidos *p*-cumaroilquínicos (ACoQ). Cada una de estas subclases se compone de al menos tres isómeros posicionales principales, con la excepción de la última clase de los compuestos de menor importancia que contiene seis isómeros principales (Clifford, Johnston, Knight, & Kuhnert, 2003; Farah, de Paulis, Trugo, & Martin, 2005). Entre estas clases, ACQ representan aproximadamente el 80 % del total contenido de ácidos clorogénicos. En particular, el 5-ACQ, el primero de estos compuestos identificados, representa casi el 60 % y por lo tanto es el isómero más estudiado y el único para el que un estándar comercial está disponible. Por esta razón, el 5-ACQ es comúnmente llamado ácido clorogénico. Los ácidos clorogénicos son responsables de la pigmentación del café y confieren astringencia, amargura y acidez a la bebida de café (Farah & Donangelo, 2006b). Sin embargo, altas cantidades en el café verde, particularmente de ACQ y AFQ, pueden producir sabor no deseable, posiblemente debido a los productos de oxidación y degradación formados antes del tostado (Farah, Monteiro, Calado, Franca, & Trugo, 2006c; Toci & Farah, 2008). Los ácidos clorogénicos son precursores de fenoles y catecoles que pueden conferir notas sensoriales desagradables que se forman durante el tostado (Trugo, 2003). El contenido de ácidos clorogénicos en *C. canephora* es generalmente dos veces mayor que en *C. arabica*, pero esta concentración varía considerablemente en ambas especies (Tabla 1.1).

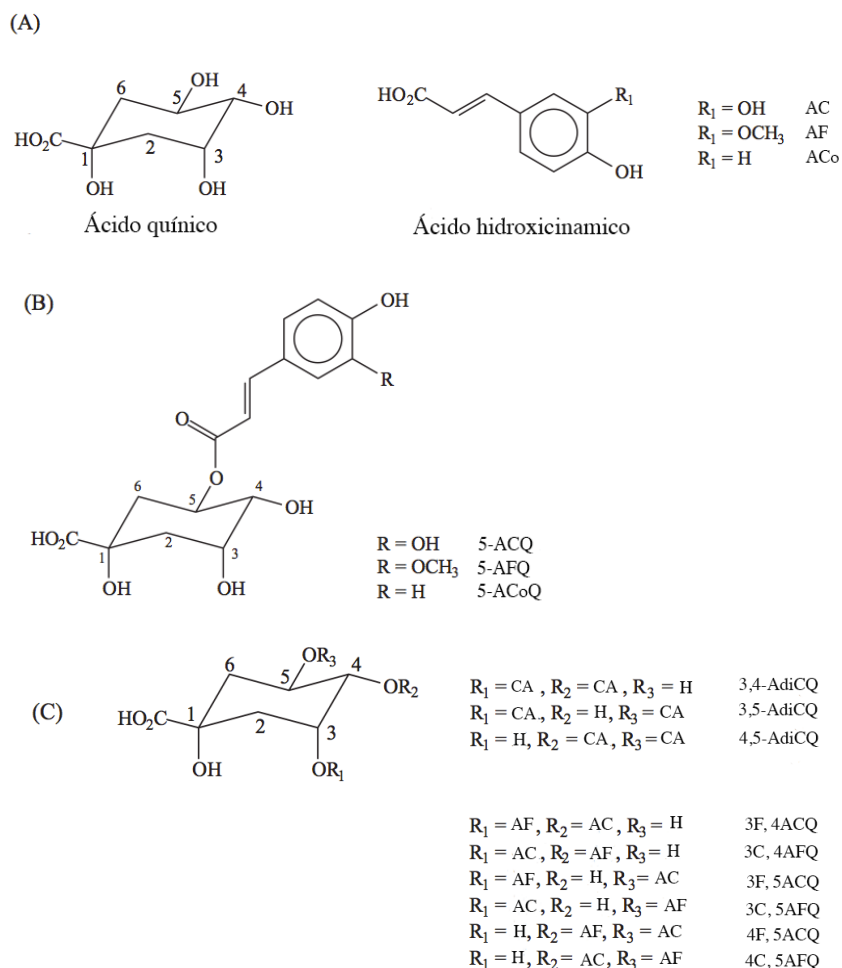


Figura 1.3. Ácidos clorogénicos y compuestos relacionados. (A) Compuestos básicos: ácido quínico y ácidos hidroxicinámicos, (B) monoésteres de ácido quínico con ácidos hidroxicinámicos (ejemplo de 5 isómeros), (C) di-ésteres de ácido quínico con ácido cafeico, y ésteres mixtos. ACQ, ácido cafeoilquínico; AdiCQ, ácido dicafeoilquínico; AFQ, ácido feruloilquínico; ACoQ, ácido p-cumaroilquínicos.

Agua

El contenido de agua en las semillas verdes de *C. arabica* y *C. canephora* varía aproximadamente 8.5-12.0 %. Por encima de este nivel, la calidad de la bebida, medida en términos de aroma y sabor, disminuye (Flament, 2002; Puertas, 2006) y la probabilidad de crecimiento microbiano aumenta. Por otro lado, una baja humedad produce grietas en las semillas y disminuye su viabilidad para germinar.

Carbohidratos

Los carbohidratos son los principales constituyentes de café y puede representar más del 50 % del peso seco. Los poli-, oligo-, di- y monosacáridos se pueden dividir en azúcares reductores y no reductores (Trugo, 1985). Los polisacáridos (solubles e insolubles) representan aproximadamente el 44 % de la materia seca en *C. arabica* y un 47 % en *C. canephora*. La sacarosa es importante para el sabor y la calidad del café. Esta representa hasta un 9 % del peso seco en los granos de *C. arabica* y aproximadamente la mitad, en granos de *C. canephora*. Pequeñas cantidades de carbohidratos simples tales como fructosa, glucosa, manosa, arabinosa, ramnosa y oligosacáridos tales como rafinosa y estaquiosa han sido identificados en el café verde (Kölling-Speer & Speer, 2005). Un contenido de sacarosa más alto es una de las razones del aroma superior y sabor general del café Arábica (Clarke, 1985).

Proteínas, péptidos y aminoácidos libres

Proteínas, péptidos, aminoácidos y ácidos libres son necesarios para la reacción de Maillard, que es la glucosilación no enzimática de proteínas y consiste en un conjunto complejo de reacciones químicas que se producen entre las proteínas y los azúcares reductores cuando se calienta el café, produciendo compuestos volátiles tales como furanos, piridinas, pirazinas, pirroles y aldehídos que son responsables del aroma del café, así como melanoidinas que son responsables del color del café y, hasta cierta medida, de su actividad antioxidante (Kölling-Speer & Speer, 2005). Los compuestos nitrogenados totales (excluyendo la cafeína y trigonelina) representan el 9-16 % de la composición química del café verde, con un contenido ligeramente más alto en *C. canephora* que en *C. arabica* (Macrae, 1985). Sin embargo, el café no es una buena fuente de nutrición de proteínas, ya que carece de los aminoácidos esenciales.

Lípidos

El contenido total de lípidos varía considerablemente entre las especies *C. arabica* y *C. canephora*. La fracción lipídica de café se compone principalmente de triacilglicerol (aproximadamente 75 %), ácidos grasos libres (1 %), esteroides (2.2 %

no esterificados y 3.2 % esterificados con ácidos grasos), y tocoferoles (0.05 %). Esta fracción también contiene diterpenos de la familia kaureno en proporciones de hasta el 20 % de la fracción total de lípidos (Kölling-Speer, & Speer, 2005; Trugo, & Macrae, 1984; Folstar, 1985). El contenido de lípido total en las semillas de café Arábica (aproximadamente 14 % de materia seca) es aproximadamente dos veces más que la de semillas café Robusta. La mayoría de los ácidos grasos en el café son insaturados. Los ácidos linoleico (18:2(*n*-6)), ácidos oleico (18:1(*n*-9)) y ácidos linoleico (18:3(3-*n*)) representan aproximadamente el 43-54 %, 7-14 %, y 1-2.6 % de la fracción de triglicéridos, respectivamente, y aproximadamente el 46 %, 11 %, y 1 % de la fracción de ácidos grasos libres, respectivamente (Nikolova-Damyanova, Velikova, & Jham, 1998; Kölling-Speer, & Speer, 2005). Las principales categorías de los esteroides en el café son 4-desmetilesteroides (representan aproximadamente 93 % del total de esteroides), 4-metilesteroides (2 %), y 4,4-dimetil-esteroides (5 %). Sitosterol pertenece a la primera categoría y representa hasta el 54 % de la fracción de esteroides; estigmasterol y campesterol representan aproximadamente el 20 % (Speer & Kölling-Speer, 2006). El contenido promedio de tocoferoles en el café ha sido reportado de 11.9 % de café verde, pero varía considerablemente dependiendo de la metodología utilizada (Ogawa, Kamiya, & Iida, 1989). Las formas α , β y γ de tocoferoles están presentes en el café. Folstar y colaboradores (1977) encontraron concentraciones de 8.9 a 18.8 % de α - tocoferol y 25 a 53 mg de β - y γ - tocoferol en aceite de café. Ogawa y colaboradores (1989) reportaron el máximo contenido de tocoferoles totales de 15.7 % en café verde, con 2.3 a 4.5 % de α - tocoferol y 3.2 a 11.4 % de β - tocoferol, el γ - tocoferol no fue detectado.

1.3 CAMBIOS EN LA COMPOSICIÓN QUÍMICA DEL CAFÉ VERDE DURANTE EL PROCESO DE TOSTADO

La composición de la semilla verde cambia dramáticamente durante el tostado como consecuencia de la pirólisis, la caramelización, y las reacciones de Maillard. Algunos

compuestos son degradados y otros se forman, incluyendo compuestos bioactivos y sustancias de alta y media volatilidad, que son importantes para el aroma y sabor de la bebida. La composición final del café tostado varía en función de la materia prima, el grado de tostado, y otras variables tales como el tipo de tostador, el tiempo, la temperatura, y la velocidad del flujo de aire en la cámara de torrefacción. En la Tabla 1.2 se muestra la composición química típica de semillas tostadas de *C. arabica* y *C. canephora*.

Tabla 1.2. Composición química de semillas tostadas de *Coffea arabica* y *Coffea canephora*.

Componente	Concentración (%)	
	<i>Coffea arabica</i>	<i>Coffea canephora</i>
<i>Carbohidratos</i>		
Sucrosa	4.2 - tr	1.6 - tr
Azúcares reductores	0.3	0.3
Polisacáridos	31 - 33	37
<i>Compuestos nitrogenados</i>		
Cafeína	1.1 - 1.3	2.4 - 2.5
Trigonelina	1.2 - 0.2	0.7 - 0.3
Proteínas/péptidos	7.5 - 10.0	7.5 - 10.0
Aminoácidos libres	ND	ND
Ácido nicotínico	0.016 - 0.026	0.014 - 0.025
<i>Lípidos</i>		
Aceite de café (triglicéridos con insaponificables, esteroides/tocoferoles)	17.0	11.0
Diterpenos (libres y esterificados)	0.9	0.2
<i>Minerales</i>		
<i>Ácidos y esteres</i>		
Ácidos clorogénicos	1.9 - 2.5	3.3 - 3.8
Ácidos alifáticos	1.6	1.6
Ácido quínico	0.8	1.0

tr: trazas; ND: No detectado.

Fuente: Clarke & Macrae (1985), Clifford (2000), Trugo & Macrae (1984), Trugo (2003), Kölling-Speer & Speer (2005), Speer & Köling-Speer (2006), Farah, de Paulis, Moreira, Trugo, & Martin (2006a), Farah & Donangelo (2006b), Holscher, Vitzthum, & Steinhart (1990), and Fischer, Reimann, Trovato, & Redgwell (2001).

Cafeína

La cafeína no se altera de manera significativa durante el tostado del café, pero pequeñas pérdidas pueden ocurrir debido a la sublimación. Sin embargo, un aumento en el contenido de cafeína se puede observar debido a la pérdida de otros compuestos.

Trigonelina

El proceso de tostado degrada a la trigonelina y la convierte en otros compuestos nitrogenados como, ácido nicotínico (3 %) y compuestos volátiles tales como pirroles (3 %), piridinas (46 %), pirazinas, y el nicotinato de metilo (Trugo & Macrae, 1985).

Ácidos clorogénicos

Debido a la inestabilidad térmica, los ácidos clorogénicos experimentan muchos cambios durante el tostado, como, isomerización, epimerización, lactonización, degradación de compuestos de bajo peso molecular (incluyendo fenoles y catecoles), y su incorporación en las melanoidinas, que contribuye al desarrollo del color y sabor. Dependiendo del grado de tueste, el contenido total de ácidos clorogénicos se reduce a menos del 1 % del contenido original. El contenido de ácidos clorogénicos en cafés tostados comerciales pueden variar desde 0.5 hasta 4 %, peso seco, dependiendo del tipo de procesamiento, mezcla, grado de tostado, método de torrefacción, y condiciones analíticas (Farah & Donangelo, 2006b; Duarte & Farah, 2009).

Agua

El contenido de humedad del café tostado (1.5-5.0 %) y varía en función del grado de tostado (Trugo & Macrae, 1985).

Carbohidratos, proteínas, péptidos y aminoácidos libres

Una porción de las proteínas del café se degradan, y los aminoácidos libres y péptidos son consumidos por la reacción de Strecker para producir aldehídos específicos con su aroma característico, CO₂ y amoniaco. Los azúcares reductores reaccionan con los aminoácidos en la conocida reacción de Maillard, formando las melanoidinas que dan

el pigmento marrón a los granos de café y otorgan sabor y color a la bebida, también se producen los pirroles, tiofenos, oxazoles, tiazoles, pirazinas, heterociclos que contribuyen con el aroma del café tostado (Flament, 2002; Murkovic & Derler, 2006). Además, las melanoidinas producidas se incorporan en las estructuras de otros componentes, tales como ácidos clorogénicos, galactomananos, y arabinogalactano-proteínas (Teixeira, Brando, Thomaziello, & Teixiera, 1995; Bekedam, Loots, Schols, Van Boekel, & Smit, 2008). La sacarosa se descompone completamente, se carameliza, produciendo pigmentos que le confieren color caramelo y sabor amargo a la bebida. La acidez de la bebida puede aumentar a medida que los niveles de ácidos alifáticos (fórmico, acético, glicólico y láctico) aumentan por la degradación de la sacarosa, polisacáridos, y otros compuestos (Clarke, 1985; Ginz, Balzer, Bradbury, & Maier, 2000).

Lípidos

La fracción de lípidos incluyendo triglicéridos y esteroides es relativamente estable al calor. Aunque los diterpenos son más sensibles al calor, cantidades razonables (0.2-0.9 % de peso seco) todavía se pueden encontrar en el café tostado, especialmente en *C. arabica*. Cafestol y kahweol son diterpenos que se degradan por el proceso de tostado. Otro diterpeno, el 16-*O*-metilcafestol, presente en el café Robusta y no en el café Arábica, se utiliza como un indicador del café Robusta en mezcla. El contenido de tocoferol también disminuye durante el tostado. Dependiendo del grado de tostado, la cantidad total de α - tocoferol, β - tocoferol, y tocoferoles totales pueden reducirse 79 a 100 %, de 84 a 100 % y de 83 a 99 %, respectivamente (Speer & Kölling-Speer, 2006). Los ácidos linoleico y palmítico son los ácidos grasos predominantes en el café.

1.4 CAFÉ DE COLOMBIA

Café de Colombia es la denominación otorgada al café 100 % Arábica (*Coffea arabica* L.) producido en la regiones cafeteras de Colombia, delimitadas entre la

latitud Norte 1° a 11°15', longitud Oeste 72° a 78°, con altitud de 400 a 2.500 metros sobre el nivel del mar (msnm) y abarca cerca de 3.3 millones de hectáreas. Este café se origina de la particular combinación de diversos factores, como latitud y altitud de la zona de cultivo de café en Colombia, sus suelos, el origen botánico de la especie y variedades de café producidas, el clima que favorece una altísima disponibilidad hídrica, la topografía cambiante, el rango de temperatura favorable en el día y durante todo el año, la luminosidad, cantidad y distribución de lluvia que generan en las plantaciones ciclos de cosechas permanentes y algunas prácticas culturales comunes que incluyen procesos de recolección selectiva y de transformación del fruto a través del proceso de beneficio por vía húmeda. Estos factores, junto con el dedicado trabajo de los productores de café, alrededor de 563.000 familias caficultoras colombianas, conducen a la producción de un café sobresaliente, suave, de taza limpia, con una acidez media/alta, cuerpo balanceado y aroma pronunciado. Es un café equilibrado con un perfil sensorial de excelente calidad (FNC, 2010a).

El mercado internacional en los últimos años ha sufrido reestructuraciones en la producción mundial de café. Colombia entre los años 2009 a 2012 descendió y pasó a ocupar el cuarto lugar. La baja en la producción se debió principalmente a problemas climatológicos (fenómeno de La Niña) pues la lluvia tiene un gran impacto en la floración y a la vez promueve enfermedades y plagas. Además, productores como Vietnam e Indonesia presentaron un crecimiento en la producción y exportación de café verde, ocupando el segundo y tercer lugar de la producción mundial, respectivamente, siendo Vietnam productor principalmente de café Robusta (Figura 1.4). Por otro lado, Perú ha presentado un crecimiento en producción y exportación en los últimos años ubicándose en el noveno lugar (ICO, 2012).

El proceso de renovación de cafetales por variedades más resistentes a enfermedades y plagas dio como resultado que Colombia sobrepasará a Indonesia nuevamente en términos de producción desde el año 2013, siendo el tercer mayor productor del

mundo y el primero en café arábigo suave. Colombia se ha consolidado como el primer productor de calidad en el mundo y compite actualmente en el segmento de café de mayor valor. Esto se debe a que se ha observado diferencias muy claras entre el café Colombiano y las demás clases de café (otras Arábicas suaves, Arábicas no lavadas y Robustas), la preferencia por el café Colombiano estimula a los productores y al país a continuar conservando esta diferencia (USDA, 2014).

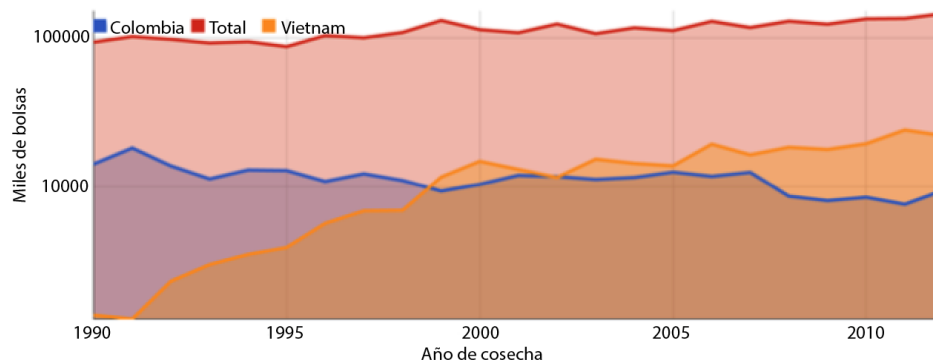


Figura 1.4. Producción mundial de café en miles de sacos x 60 kg y por año de cosecha (1990 - 2012), para los países de Colombia y Vietnam. Fuente: ICO, 2012.

Colombia goza de un reconocimiento internacional como productor de café Arábica suave de alta calidad, facilitando en gran medida la labor de promoción de “café especiales” o “de origen” por parte de la Federación Nacional de Cafeteros de Colombia (FNC). Para garantizar la autenticidad y origen del café de Colombia se requiere de diversos instrumentos legales, como la Indicación Geográfica Protegida (IGP) otorgada por la Comisión de la Unión Europea en octubre de 2007, como distinción y reconocimiento a su calidad y reputación. Los productos con una IGP son producidos, procesados o elaborados en una zona geográfica específica que toma su nombre y poseen una reputación, característica o cierta cualidad atribuible a esa zona geográfica. La Denominación de Origen Protegida (DOP) es otorgada por ciertos países de Sudamérica. Los productos con una DOP son producidos, procesados y elaborados dentro de una zona geográfica determinada, y cuya calidad se encuentra vinculada directamente con ese origen. La Superintendencia de Industria y Comercio expidió la Resolución No. 4819 por medio de la cual declara la protección de la DOP

Café de Colombia por solicitud de la Federación Nacional de Cafeteros, en marzo de 2005. Norteamérica, reconoce al café de Colombia a través de la Marca de Certificación (MC), con la cual se garantiza que el café cumple con unos estándares de calidad previamente definidos en normas dadas (FNC, 2010b).

El café verde de exportación debe pasar por diferentes puntos de análisis de calidad sensorial, granulometría y humedad, como se ha establecido en las regulaciones vigentes en Colombia. Quien controla y expide el certificado de calidad es la empresa Almacenes Generales de Depósito de Café - ALMACAFE S.A., la cual cuenta con autonomía administrativa y patrimonio de la FNC contribuyendo a proteger la calidad del Café de Colombia a lo largo de todo el proceso, desde su recibo en las diferentes regiones colombianas hasta su embarque a los mercados externos, o su venta a la industria nacional (Gallego, 2007). Por otra parte, Colombia tiene que importar café de menor calidad de los países vecinos, para abastecer su demanda interna, ya que el 95 % de la producción es exportada por los mejores precios en el exterior, generando condiciones para fraudes del café Colombiano. En este contexto, es importante asegurar la calidad del origen botánico y geográfico de los granos de café, aprovechando los avances en la ciencia y tecnología, como el uso de plataformas analíticas combinadas con modelos quimiométricos que se puedan aplicar idealmente al material verde o tostado y a una gran cantidad de muestras (FNC, 2010c).

1.5 PLATAFORMAS ANALÍTICAS

Actualmente, dada la calidad del café 100 % Arábica, este es el preferido por el consumidor. Ello se refleja en el precio de venta que suele ser un 20-25 % más alto que el de café Robusta o mezclas. Si bien la declaración del origen geográfico del café le otorga un valor agregado al producto colombiano, con la globalización de los mercados de alimentos y la facilidad con que los alimentos son transportados en todo el mundo, aumenta las preocupaciones del consumidor sobre el origen del alimento y si realmente está pagando por el café de alta calidad que se oferta (Kelly *et al.*, 2005).

El café puede ser objeto de fraude por medio de adulteraciones con productos más baratos o falsas denominaciones en relación a su especie u origen geográfico, afectando su calidad, reputación y precio final. En los últimos años ha habido un incremento en la falsificación de la declaración de los productos alimenticios, por lo tanto, es importante contar con herramientas que permitan asegurar la calidad, procesamiento y seguridad de materias primas y productos finales de los alimentos.

Una gran variedad de técnicas analíticas han mostrado ser prometedoras para detectar cafés adulterados, como espectrometría de masas de relaciones isotópicas (IRMS, por su sigla en inglés, Isotope Ratio Mass Spectrometers), espectrometría de masas con ionización por electrospray (ESI-MS, por su sigla en inglés, Electrospray Ionization-Mass Spectrometry), cromatografía de gases acoplada a espectrometría de masas (GC-MS, por su sigla en inglés, Gas Chromatography–Mass Spectrometry), cromatografía líquida de alta eficacia (HPLC, por su sigla en inglés, High Performance Liquid Chromatography), espectroscopia Raman, espectrometría de masas con plasma inductivamente acoplado (ICP-MS, por su sigla en inglés, Inductively Coupled Plasma Mass Spectrometry), espectroscopia de infrarrojo cercano (NIRS, por su sigla en inglés, Near-Infrared Spectroscopy) y resonancia magnética nuclear (RMN).

En la Tabla 1.3 se presenta un resumen de los estudios metabolómicos realizados en los últimos 20 años con diferentes propósitos en los que combinaron diferentes técnicas analíticas con quimiometría para el análisis de compuestos químicos del café, como cafeína, trigonelina, ácidos clorogénicos, lípidos, carbohidratos, entre otros.

Tabla 1.3 Enfoques metabolómicos para evaluar la autenticidad del café publicados en los últimos 20 años. (*n*= Número de muestras, T= tostado y V= verde)

Propósito del análisis	Técnica analítica	Presentación del café	<i>n</i>	Análisis estadístico	Compuestos químicos del café	Referencias
Adulteración	FT-IR	T	146	PCA, LDA, ANN	Carbohidratos	Briandet, Kemsley & Wilson, 1996a
	HPLC	T	21	PLS	Tocoferol	Jham <i>et al.</i> , 2007
	GC-MS	T	12	PCA	Volátiles	Oliveira <i>et al.</i> , 2009
	NIRS	T	158	PLS, PLS-GA		Ebrahimi-Najafabadi <i>et al.</i> , 2012
	Direct Infusion MS	T	13	PCA		Garrett <i>et al.</i> , 2012
Autenticación 3 productores	DRIFTS	T	165	PCA, LDA		Reis, Franca & Oliveira, 2013
	HPLC	T	10	PCA, LDA	Carbohidratos	Pauli <i>et al.</i> , 2014
	¹ H-NMR	T	98	PCA, LDA, SIMCA		Charlton, Farrington & Brereton, 2002
	FTIR	T	17	PLS, PCR		Wang <i>et al.</i> , 2009
	NIRS	T	82	PLS-DA	Ácidos clorogénicos, sacarosa, lípidos	Santos <i>et al.</i> , 2012
Autenticación de marcas	GC-FID	T	27	PCA	Kahweol, 16-OMC	Pacetti <i>et al.</i> , 2013
Autenticación de mezclas	GC-MS	T	21	PCA, OPLS-DA	Ácidos cítrico, málico y relación de inositol/ácido piroglutámico	Jumhawan <i>et al.</i> , 2013
Autenticación café Kopi Luwak		T				
Autenticación de café orgánico	HS PTR-MS/	T	110	PLS-DA	Volátiles	Özdestan <i>et al.</i> , 2013
	PTR-TOF-MS	T				
	NIRS	T	82	PLS-DA	Ácidos clorogénicos, sacarosa, lípidos	Sarraguça <i>et al.</i> , 2013
	UV-Vis	T	175	SIMCA, SPA-LDA		Souto <i>et al.</i> , 2010
	NIRS	T	48	ANOVA, PLS	Pérdida de peso, densidad, humedad	Alessandrini <i>et al.</i> , 2008
Origen botánico (<i>C. arabica</i> vs <i>C. canephora</i>)	FT-IR	V	28	PCA, DA		Kemsley, Ruault & Wilson, 1995
	HPLC	V	20		Ácidos hidroxycinnámicos	Andrade <i>et al.</i> , 1998
	GC-FID	V	31	PCA, CA	Lípidos, esteroides	Carrera <i>et al.</i> , 1998
	HPLC	V	41	PCA, CA, KNN	Ácidos clorogénicos, cafeína, trigonelina	Martin <i>et al.</i> , 1998a
					aminoácidos libres totales, polifenoles	
Origen botánico (<i>var. Sarchimor</i> y <i>Catuai</i>)	ICP-AES	V	41	PCA, CA	Multi-elementos (11)	Martin, Pablos & González, 1998b
	GC-FID/GC-MS	V y T	32	PCA, HCA	Volátiles	Costa Freitas & Mosca, 1999
	HPLC	T	29	PCA, HCA	Trigonelina, Ácido nicotínico, cafeína	Casal <i>et al.</i> , 2000
	HPLC	V y T	32	PCA, LDA	Tocoferoles, triglicéridos	González <i>et al.</i> , 2001
	GC-FID	V	30	HCA, CDA	Aminoácidos libres	Casal <i>et al.</i> , 2003
	GC-FID	V y T	24	CA, CVA, DA	Ácidos grasos	Alves, Casal, Oliveira & Ferreira, 2003
	NIRS	T	83	PCA	Cafeína, ácidos clorogénicos, ácidos total	Esteban-Díez <i>et al.</i> , 2004
	RAMAN	V y T	168	PCA	Kahweol, cafestol	Rubayiza & Meurens, 2005
	ESI-MS	V	20	PCA, HCA	Compuestos fenólicos	Mendonça <i>et al.</i> , 2008
	HPLC	V y T	29	ANOVA, CA	Tocoferoles	Alves, Casal, Alves & Oliveira, 2009
	GC-MS	T	30	PLS-DA	Furanos, fenoles, piridinas, alcoholes, compuestos aromáticos	Hovell, Pereira, Arruda & Rezende, 2010
	RAMAN	V	32	PCA	Ácidos clorogénicos, lípidos	El-Abassy, Donfack & Materny, 2011
	ESI FT-ICR-MS	V	23	PCA, PLS-DA	Compuestos fenólicos, ácidos grasos, sacarosa, glicósidos de diterpeno	Garrett <i>et al.</i> , 2013
					Ácidos clorogénicos, cafeína	
Origen botánico (mezclas entre <i>C. arabica</i> y <i>C. canephora</i>)	FT-IR	T	70	PCA, PLS		Briandet, Kemsley & Wilson, 1996b
	NIR-MIR	T	56	FDA, PLS		Downey <i>et al.</i> , 1997
	ICP-AES	T	30	PLS	Multi-elementos (11)	Martin, Pablos & González, 1999
	GC-FID	T	13	PCR	Lípidos, esteroides	Valdenebro <i>et al.</i> , 1999
	NIRS	T	191	PCA		Esteban-Díez <i>et al.</i> , 2007
	NIRS	T	191	PLS		Pizarro <i>et al.</i> , 2007
	¹ H-NMR	T	22	OPLS		Cagliani <i>et al.</i> , 2013
	GC-FID	T	41	PCA, PLS	Ácidos grasos	Romano <i>et al.</i> , 2014
	ICP-AES	T	160	PCA, CDA, DFA, ANN	Multi-elementos (19)	Anderson & Smith, 2002
Origen geográfico						
	EA-P-IRMS	V	45	LDA, CART	Relaciones isotópicas de C, H y O	Weckerle <i>et al.</i> , 2002
	IRMS	V	46	PCA	Relaciones isotópicas de C, N y B	Serra <i>et al.</i> , 2005
	NIRS	V	120	PCA	Humedad, lípidos, proteínas, sacarosa, cafeína, trigonelina, ácidos orgánicos, ácidos clorogénicos	Haiduc, Gancel & Ieloup, 2006
	GC-TOF-MS	T	47	PCA	Volátiles y semi-volátiles	Risticvic, Carasek & Pawliszyn, 2008
	IRMS/EA	V	68	PCA	Relaciones isotópicas de C, N y O, composición porcentual de C y N	Rodrigues <i>et al.</i> , 2009
	HPLC	V	107	PCA, LDA, PLS-DA, CART	Ácidos clorogénicos, cinamoil amidas, cinamoil glucósidos, ácidos fenólicos	Alonso-Salces <i>et al.</i> , 2009
	LC-MS/GC-FID	T	21	PCA	Proteínas totales, carbohidratos totales, monosacáridos, aminos	Choi <i>et al.</i> , 2010
	IRMS/MC-ICP-SFMS	V	47	CDA	Relaciones isotópicas de C, N, S, O y Sr, multi-elementos (30)	Rodrigues <i>et al.</i> , 2011a
	MC-ICP-MS/IRMS	V	60	PCA	Relaciones isotópicas de Sr y O	Rodrigues <i>et al.</i> , 2011b
	ICP-MS/IRMS	V	64	CDA	Relaciones isotópicas de C, H, N y O, multi-elementos (54)	Santato <i>et al.</i> , 2012
Origen geográfico	¹ H-NMR	T	40	OPLS-DA		Consonni, Cagliani & Cogliati, 2012
	¹³ C-NMR	V	60	PCA, OPLS-DA	Sacarosa, cafeína, ácidos clorogénicos, colina, AA, ácidos orgánicos, trigonelina	Wei <i>et al.</i> , 2012a
	ICP-MS/ICP-AES	V y T	42	LDA, PCA	Multi-elementos (59)	Jena & Watling, 2013
	FT-IR	V	18	ANN, SIMCA, MLP		Link <i>et al.</i> , 2014
	MC-ICP-MS	V	21	PCA	Relaciones isotópicas de B y Sr, multi-elementos (7)	Liu <i>et al.</i> , 2014

Como se puede apreciar alrededor del 77 % de estas investigaciones presentan un número de muestras menor a 80, algunos autores aclaran que el número de muestras de café usadas es relativamente pequeño para la generalización y se tiene una posibilidad de clasificación equivocada. Por lo tanto, recomiendan adición de muestras para la aplicación del enfoque integrado de metabolómica y así reducir al mínimo la posibilidad de resultados falsos (Choi *et al.*, 2010; Liu *et al.*, 2014).

Liu y colaboradores (2014) basados en evaluaciones de múltiples trazadores isotópicos, han encontrado que las principales dificultades son debidas a la limitación de la discriminación completa utilizando un único enfoque. Comparando el análisis multi-relación isotópica y análisis multi-elemento, este último parece tener más ventaja cuando se combina con técnicas quimiométricas, pero los mejores resultados posibles proceden de la combinación de las dos técnicas. Sin embargo, la implementación de esta estrategia es bastante difícil en el análisis de rutina (Drivelos & Georgiou, 2012).

1.6 METABOLÓMICA BASADA EN RMN

Desde una perspectiva de la química analítica, los alimentos pueden ser considerados matrices complejas sólidas, semi-sólidas o líquidas que contienen cientos de miles de compuestos (nutrientes o no): proteínas, péptidos, lípidos, carbohidratos y muchos otros compuestos pequeños, tales como aminoácidos, ácidos orgánicos, ácidos nucleicos, ácidos grasos, fitoquímicos y minerales. La metabolómica se centra en el análisis de los metabolitos de molécula pequeña en los sistemas biológicos (Fiehn, 2002) y consiste en una secuencia de pasos que incluye la preparación de la muestra, extracción de los metabolitos, detección (técnicas analíticas) y tratamiento de los datos (quimiometría).

La aplicación de la metabolómica ofrece enormes oportunidades para obtener información detallada que se puede correlacionar directamente a la calidad, seguridad

y otras características relacionadas con el procesamiento, almacenamiento y evaluación de autenticidad de los alimentos (Cevallos-Cevallos *et al.*, 2009). Dos enfoques analíticos diferentes se siguen en la metabolómica no dirigida: a) Perfil metabolómico se refiere al análisis (identificación y cuantificación) de una clase de metabolitos predefinidos (metabolitos químicamente relacionados o relacionados con una vía metabólica en particular); y, b) Huella metabolómica se refiere al análisis del conjunto total de metabolitos, evitando prejuicios contra ciertas clases de compuestos, para la clasificación rápida de muestras. Este análisis ofrece una clasificación sin identificar y cuantificar metabolitos individuales. Ambos enfoques metabolómicos se utilizan en la búsqueda de nuevos biomarcadores de calidad y autenticidad, y, sin duda, contribuyen al conocimiento bioquímico de los alimentos y la comprensión de los mecanismos de acción en los procesos de producción de los mismos (Dunn & Ellis, 2005; Ibáñez *et al.*, 2013).

Actualmente hay dos herramientas principales utilizadas para la investigación metabolómica, métodos cromatográficos acoplados a espectrometría de masas y RMN. Comparando entre las dos técnicas, la RMN tiene una desventaja en términos de sensibilidad absoluta, pero tiene claras ventajas como (Wishart, 2008):

- Capacidad para determinar diversas clases de compuestos químicos de una manera rápida y simultánea.
- Proporciona información cualitativa y cuantitativa de los compuestos químicos.
- Requiere mínima preparación de la muestra.
- Tiempo de análisis más corto.
- Naturaleza no destructiva.
- Buena reproducibilidad.

Se puede compensar la sensibilidad intrínseca de la RMN mediante la adquisición y suma de más decaimientos (FID, por su sigla en inglés, Free Induction Decay), usando campos más altos o sondas criogénicas que permitan mejorar la razón señal a ruido.

En el análisis metabolómico basado en RMN se deben superar problemas como las amplias resonancias que producen las macromoléculas debidas a la difusión rotacional limitada y tiempos de relajación T_2 cortos, provocando dificultades en la interpretación espectral, para tal fin se emplean las secuencias de pulsos 1D-NOESY-Presat (McKay, 2011) y 1D-CPMG (Carr-Purcell-Meiboom-Gill) (Meiboom & Gill, 1958). 1D-NOESY se ha convertido en la secuencia más popular para el análisis metabolómico basado en RMN. Esto se debe principalmente a la alta calidad de supresión de agua con poca calibración y elevada coherencia en los espectros obtenidos. CPMG, se utiliza para eliminar las amplias señales de proteínas, que aparecen en el espectro si es que estas macromoléculas no han sido removidas del extracto, antes de la medición de RMN.

Presaturación, es un simple experimento de dos pulsos. El primero, “*Presaturación*” es relativamente largo (1-3 segundos) y de baja potencia de radiofrecuencia (decenas de Hz), sirve para saturar selectivamente una frecuencia específica (generalmente la frecuencia del agua). El segundo, “*Excitación*” de 45-90° no selectivo, es usado para excitar las resonancias deseadas, como se aprecia en la Figura 1.5.

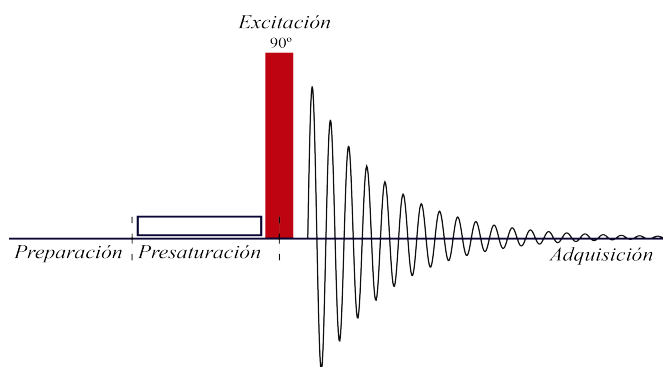


Figura 1.5. Secuencia de pulso de Presaturación-Excitación. Preparación es el tiempo de recuperación para restablecer el equilibrio, presaturación es representado con una barra horizontal que indica el largo periodo de saturación a baja potencia, excitación es representado con una barra roja que indica un pulso de 90° y la adquisición de la señal de RMN se muestra por un decaimiento de inducción libre.

Esta secuencia de pulsos es particularmente útil para muestras acuosas o aquellas con una señal de disolvente grande. Con la optimización adecuada, el espectro resultante puede estar libre de la señal del disolvente y dar lugar a una mejor relación señal-ruido de las resonancias del soluto debido a la reducción en el rango dinámico y el consiguiente aumento de la ganancia disponible.

La secuencia de pulsos 1D-NOESY-Presat de la Figura 1.6 inicia con un largo periodo de saturación a baja potencia “*Presaturación*” permitiendo la selectividad de la resonancia del disolvente, seguido de dos pulsos de 90° que generan inversión en el estado de equilibrio de los espines. Durante el tiempo de mezcla “*Mezcla*” la relajación cruzada conduce a una transferencia de magnetización incoherente a lo largo del eje longitudinal, el último pulso de 90° envía la magnetización longitudinal al plano transversal, haciendo que sea observable durante el periodo de adquisición, T_2 (McKay, 2011). En general, en macromoléculas se emplean tiempos de mezcla menores de 200 ms, para evitar los efectos de la difusión de espín, mientras que en moléculas pequeñas pueden emplearse tiempos de mezcla de varios centenares de ms, hasta llegar a 1 s. Es importante realizar un control fino de la temperatura, el número de scans y el incremento óptimo para conseguir las señales adecuadas con el menor ruido posible.

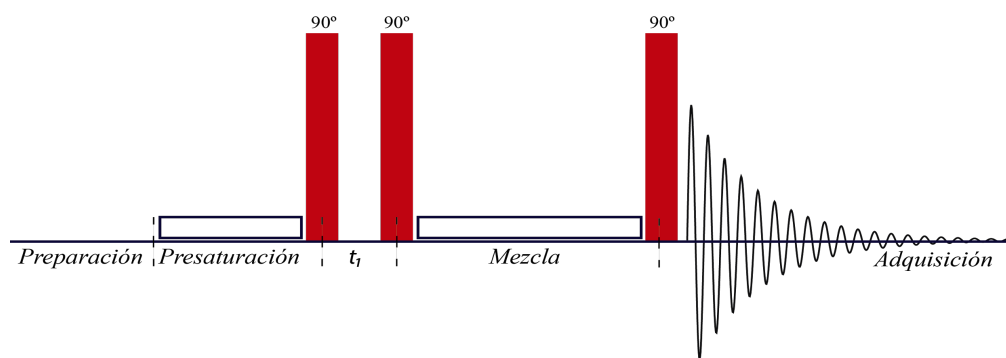


Figura 1.6. Secuencia de pulsos de 1D-NOESY-Presat. Preparación es el tiempo de recuperación para restablecer el equilibrio, presaturación y mezcla son representados por las barras horizontales que indican los largos periodos de saturación a baja potencia, las barras rojas representan pulsos de 90° y la adquisición de la señal de RMN se muestra por un decaimiento de inducción libre.

Las moléculas de alto peso molecular, como proteínas y lipoproteínas, son de tamaño mayor, se reorientan más lentamente en solución (*tumbling*) y por tanto experimentan una relajación transversal más rápida. Este fenómeno se utiliza en metabolómica como un filtro para eliminar las amplias resonancias asociadas con macromoléculas de alto peso molecular que distorsionan la línea base, facilitando la observación y resolución de las señales correspondientes a metabolitos de bajo peso molecular. La secuencia de pulsos 1D-CPMG con presaturación (Figura 1.7) permite la edición del tiempo de relajación transversal T_2 , en la cual un primer pulso de 90° es aplicado durante un tiempo suficiente para voltear la magnetización del espín en el plano transversal. Los espines precesan al unísono, produciendo una gran señal, pero luego se desfasan rápidamente debido a las inhomogeneidades del campo magnético. Luego se aplica otro pulso de 180° durante un tiempo suficiente como para invertir su dirección de precesión, haciendo que vuelvan a estar en fase después de un corto tiempo. Cuando los espines están en fase producen otra señal intensa denominada eco, pero pronto se desfasan nuevamente y vuelven a ser puestos en fase con otro pulso de 180° . Este proceso de refocalización se repite muchas veces, a la vez que se mide la magnitud de cada eco, la cual se reduce con el tiempo debido a los mecanismos de relajación molecular. Una medición puede incluir muchos cientos de ecos, en tanto que el tiempo entre cada eco (el espaciamiento entre ecos) es del orden de 1 ms o un valor inferior (Carr & Purcell, 1954; Meiboom & Gill, 1958; Zhang & Hirasaki, 2003).

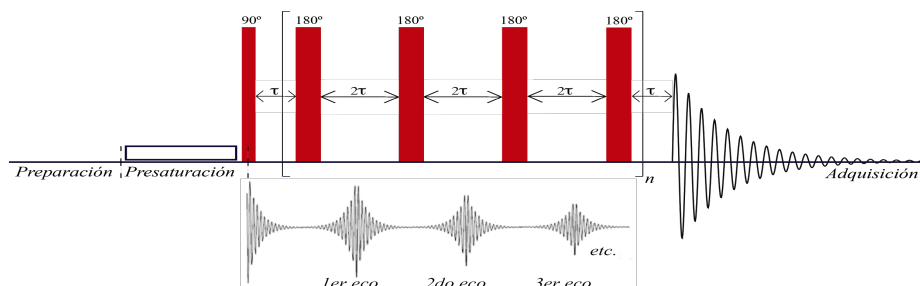


Figura 1.7. Secuencia de pulsos de 1D-CPMG con presaturación. Un pulso de excitación de 90° y varios pulsos de 180° para refocalización. Después de cada uno de los pulsos de 180° aparecen ecos. El tiempo de relajación transversal T_2 se determina a partir del decaimiento de la señal.

1.7 PROCESAMIENTO DE LOS DATOS

Los datos obtenidos por RMN a menudo necesitan ser procesados con el fin de ajustarse a los requisitos previos para el análisis estadístico de datos multivariados y producir resultados fiables e interpretables. El procesamiento permite transformar los datos obtenidos, señal FID, en espectros mediante tratamientos matemáticos, como transformada de Fourier, apodización, corrección de la línea de base y corrección de fase (Goodacre *et al.*, 2007). La apodización consiste en multiplicar la señal FID por funciones envolventes, permitiendo dar más peso a una u otra región de la señal FID reduciendo el ruido, aumentando la relación señal/ruido. Otro proceso importante es el *binning* o *bucketing*, que consiste en la subdivisión e integración del espectro, en pequeñas regiones denominadas *bins* o *buckets*. Este proceso permite obtener la integración de la intensidad de la señal dentro de cada *bin* o *bucket* para producir un conjunto más pequeño de variables (Worley & Powers, 2013). El procesamiento de los datos también incluye la exclusión de regiones espectrales con información que no es reproducible entre muestras o que no corresponde a señales de metabolitos, como la región del solvente.

1.8 ANÁLISIS ESTADÍSTICO MULTIVARIANTE

La mayoría de los estudios metabolómicos resultan en conjuntos de datos multivariantes altamente complejos y variables, que necesitan la aplicación de métodos matemáticos y estadísticos para explorar y extraer información más útil del conjunto de datos medidos. Existen diferentes categorías de análisis estadístico multivariado:

- a. Análisis exploratorio de datos o métodos de análisis no supervisados se utilizan a menudo para determinar si existe alguna agrupación intrínseca dentro de un conjunto de datos, sin un conocimiento a priori de la clase de muestra, mediante la reducción de la información a una forma más comprensible y para exponer posibles valores atípicos. El análisis de componentes principales (PCA, por su

sigla en inglés, Principal Component Analysis) y el análisis de agrupamiento jerárquico (HCA, por su sigla en inglés, Hierarchical Cluster Analysis) son algunos de los algoritmos de exploración.

- b. La regresión lineal se utiliza para desarrollar modelos de calibración que correlacionan la información de un conjunto de mediciones conocidas a la propiedad de interés. Mínimos cuadrados parciales (PLS, por su sigla en inglés, Partial Least Squares) y la regresión de componentes principales (PCR, por su sigla en inglés, Principal Components Regression) son dos algoritmos comúnmente utilizados para la regresión y están diseñados para evitar los problemas asociados con el ruido en los datos.
- c. Modelos de clasificación o métodos de análisis supervisados se aplican en situaciones donde se requieren clasificar muestras en categorías predefinidas o clases. Un modelo de clasificación se utiliza para asignar la clase de una muestra mediante la comparación de esa muestra en un conjunto de datos analizados anteriormente, para los que sus categorías son ya conocidos. PLS discriminant analysis (PLS-DA, por su sigla en inglés, Partial Least Squares Discriminant Analysis), orthogonal PLS discriminant analysis (OPLS-DA, por su sigla en inglés, Orthogonal Partial Least Squares Discriminant Analysis) y linear discriminant analysis (LDA, por su sigla en inglés, Linear Discriminant Analysis) son algunos de los más empleados (Jove *et al.*, 2011).

El PCA es un procedimiento matemático ampliamente utilizado para transformar un conjunto de variables originales correlacionadas en un nuevo conjunto de variables subyacentes no correlacionadas (ortogonales), conocidas como componentes principales (PCs), utilizando la matriz de covarianza. Las nuevas variables son combinaciones lineales de las originales. Los PCs reducen los datos de una manera que maximiza la variación espectral entre muestras, así un gran conjunto de datos puede ser representado por sólo unos pocos PCs, que describen una gran parte de la varianza en los datos como una combinación lineal de las variables originales. (Halouska & Powers, 2006). PCA trabaja descomponiendo la matriz X (que contiene

el conjunto de datos originales) como el producto de dos matrices más pequeñas, que se llaman matriz de carga y de puntuación, más una matriz de residuos E . La matriz de carga (P) contiene información sobre las variables, se compone de un par de vectores (PCs) que son combinaciones lineales de las X -variables originales. La matriz de puntuación (T) contiene información acerca de los objetos. Cada objeto se describe en términos de sus proyecciones sobre los PCs (en lugar de las variables originales) (Ecuación 1.1).

$$X = TP^T + E \quad \text{Ec. 1.1}$$

Por otro lado, PLS es un método de regresión multivariable que considera dos grupos de variables, independientes (matriz X) y dependientes (matriz Y), permitiendo correlacionar los dos conjuntos de datos entre sí, o la predicción de un valor de un conjunto mediante el uso de los datos desde el otro conjunto que es más fácil de medir. PLS puede manejar más de una variable dependiente y no está influenciada significativamente por la correlación entre las variables independientes. En el modelo PLS se trata de encontrar el sentido multidimensional en el espacio de X que explica la dirección de la máxima varianza multidimensional en el espacio Y (Wold *et al*, 2001a).

La regresión PLS consta de dos procesos, inicialmente se transforma la matriz de predictores X de orden $n \times p$, con ayuda del vector de respuestas Y de orden $n \times 1$, en una matriz de componentes o variables latentes no correlacionados, $T = (T_1, \dots, T_p)$ de orden $n \times p$, llamados componentes PLS; esto contrasta con el análisis de componentes principales en el cual los componentes son obtenidos usando sólo la matriz de predictores X . Luego, se calcula el modelo de regresión estimado usando el vector de respuestas original y la matriz de componentes PLS (predictores). La reducción de la dimensionalidad puede ser aplicada directamente sobre los componentes ya que estos son ortogonales. El número de componentes necesarios

para el análisis de regresión debe ser mucho menor que el número de predictores (Wold *et al*, 2001b).

PLS-DA consiste en una regresión PLS clásica donde las variables de respuesta representan grupos o clases. Los valores de las variables explicativas (independientes) y de respuesta (dependientes) se registran en la matrix X de dimensión $n \times m$ (Ecuación 1.1) y en la matriz Y de dimensión $n \times p$ (Ecuación 1.2), respectivamente. Los datos de las filas de la matrix X representan n objetos con m características (variables explicativas), y Y describe para el mismo n objetos p propiedades (variables de respuesta categóricas), las cuales son variables binarias que describen las diferentes categorías, por ejemplo, -1 y 1 en el caso de dos categorías (Barker & Rayens, 2003). El número de variables dependientes es igual al número de categorías. El algoritmo está optimizado para el caso equilibrado, lo que significa que hay la misma cantidad de miembros en cada categoría (Pérez-Enciso & Tenenhaus, 2003). El modelo PLS-DA puede ser expresado por:

$$\text{Modelo } X = TP^T + E \quad \text{Ec. 1.1}$$

$$\text{Modelo } Y = UQ^T + F \quad \text{Ec. 1.2}$$

donde, X es la matrix de predictores ($n \times m$), Y es la matrix de respuesta ($n \times p$), T y U son las matrices de proyecciones ($n \times l$) de X (puntuaciones X) y Y (puntuaciones Y), respectivamente, P y Q son las matrices de carga ($m \times l$) y ($p \times l$), respectivamente, y las matrices E y F son los términos de error.

OPLS-DA es un método de visualización de datos que es útil para la observación de grupos dentro de los datos multivariados. Este análisis supervisado utiliza la regresión lineal, donde la pertenencia a la clase de muestras se incluyen en el cálculo. OPLS-DA como una extensión de PLS-DA con un OSC-filtro integrado puede eliminar la variabilidad que no es relevante para la separación de las clases. OPLS es capaz de girar la proyección de manera que el modelo se centra en el efecto de interés, lo que

permite una mejor distinción entre grupos (Bylesjö *et al.*, 2006). OPLS-DA utiliza la información en la matriz de respuesta categórica Y para descomponer la matriz X en tres partes distintas, variación correlacionada (predictiva), variación no correlacionada (ortogonal) y residuos del modelo, como se describe en la Ecuación (1.3).

$$\text{Modelo } X = T_p P_p^T + T_o P_o^T + E \quad \text{Ec. 1.3}$$

Donde, T y P son las matrices de puntuaciones y cargas para X , respectivamente, y los subíndices indican si explican la variación predictiva (p) u ortogonal (o). E denota la matriz residual de X (Trygg & Wold, 2002).

Un t -test multivariante (T^2 de Hotelling) se utiliza para asignar una probabilidad de que el valor de la puntuación prevista de alguna observación concreta es diferente de la media de la distribución del conjunto de observaciones en las matrices de puntuaciones obtenidas por PCA, PLS-DA y OPLS-DA, con un intervalo de confianza del 95 % (Hotelling, 1931).

En los métodos supervisados es muy importante evaluar la calidad de los análisis realizados y la fiabilidad de los modelos obtenidos, ya que estos modelos pueden estar propensos a sobreajuste (Westerhuis *et al.*, 2008). Por lo tanto, la técnica de validación cruzada (VC) se adopta para evaluar la capacidad predictiva de los modelos y seleccionar el número óptimo de componentes o variables latentes. Esta forma de validación significa que el conjunto de datos se divide aleatoriamente en dos subconjuntos, un subconjunto de datos de entrenamiento o calibración para construir el modelo y un subconjunto de datos de prueba para evaluar la capacidad de predicción del modelo, donde el subconjunto de prueba en ningún modo es utilizado para generar el modelo entrenado (Broadhurst & Kell, 2006). Este proceso se repite n veces de forma que todas las muestras han estado en la construcción y validación del modelo. Para cada uno de estos n modelos entrenados se obtiene una predicción del

conjunto de validación y los errores de predicción de todos los modelos son almacenados, el resultado final de los errores se obtiene a partir de la media aritmética de los n valores de errores obtenidos. Un proceso más robusto para evaluar la calidad de la capacidad predictiva del modelo es la doble validación cruzada (n VC) que consiste en dos bucles anidados de validación cruzada que frecuentemente son referenciados como bucle de validación interna y externa. En la Figura 1.8 se presenta simplificado el proceso seguido. En el bucle externo de doble validación cruzada todos los datos son divididos al azar en dos subconjuntos, correspondientes a un conjunto de datos para la construcción del modelo y otro conjunto de datos de validación externa que será utilizado exclusivamente para evaluar el modelo. Posteriormente, el conjunto de datos para la construcción del modelo se utiliza en el bucle interno de doble validación cruzada, el cual sigue el mismo proceso de una validación cruzada (Smit *et al.*, 2007; Westerhuis *et al.*, 2008; Szymańska *et al.*, 2012).

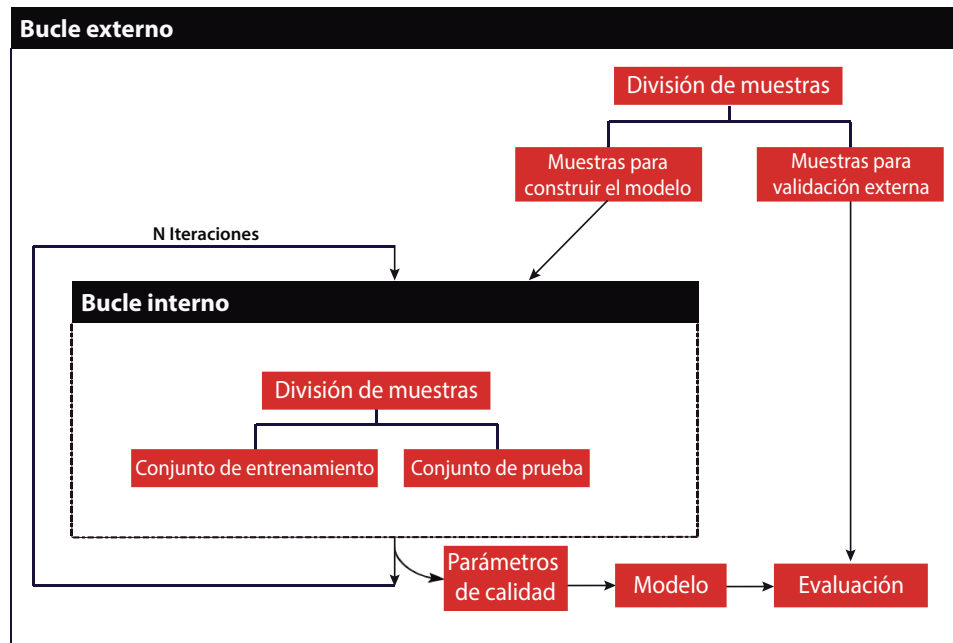


Figura 1.8. Esquema simplificado de doble validación cruzada. Comprende dos bucles, un bucle interno para la estimación del número óptimo de variables a utilizar en la clasificación y otro externo para la evaluación del modelo seleccionado.

Para evaluar la capacidad predictiva del modelo se suele usar la suma residual de cuadrados del error de predicción (*PRESS*, por su sigla en inglés, predicted residual sum of squares) (Ecuación 1.4) o el error cuadrático medio (*MSE*, por su sigla en inglés, Mean Squared Error), que se obtiene dividiendo *PRESS* por el número de muestras, (Ecuación 1.5). Se puede calcular el MSE de la calibración (*MSEC*), la predicción de la validación interna (*MSECV*) y la predicción de la validación externa (*MSEEP*).

$$PRESS = \sum_i (y_i - \hat{y}_i)^2 \quad Ec. 1.4$$

$$MSE = \frac{1}{z} \sum_{i=1}^z (y_i - \hat{y}_i)^2 \quad Ec. 1.5$$

donde, $\hat{y}$ se refiere al valor predicho de pertenencia a una clase para la muestra i mientras z es el número de muestras.

Determinar la cantidad correcta de componentes o variables latentes consiste en calcular el valor de *MSECV* frente al número de PCs y buscar el mínimo. Al aumentar el número de PCs disminuye el error, debido a que se modela cada vez mejor el sistema, hasta que llega un momento en que los nuevos PCs introducidos explican ruido y *MSECV* vuelve a aumentar por sobreajuste del modelo. Pero en el caso que se tenga un conjunto de predicción o validación externo, el número de PCs se selecciona en función de la evolución del error para este conjunto de muestras.

El uso de la validación cruzada da la posibilidad de calcular las estadísticas de los modelos, como R^2 que es la proporción de varianza explicada y Q^2 que es la proporción de la varianza predicha. El Q^2 está basado en la evaluación del error entre la variable categórica predicha $\hat{y}$, y la conocida y (Ecuación 1.6). *TSS* es la suma total de cuadrados de la diferencia de la variable dependiente y su media general (por su sigla en ingles, Total Sum of Square) (Ecuación 1.7) (Cruciani *et al.*, 1992):

$$Q^2 = 1 - \frac{PRESS}{TSS} \quad Ec. 1.6$$

$$TSS = \sum_i (y_i - \bar{y}_i)^2 \quad Ec. 1.7$$

donde, $\bar{y}$ se refiere al valor medio de y para todas las muestras.

El valor óptimo de 1 para Q^2 es difícil de alcanzar, ya que esto requiere que la predicción de la clase de cada individuo debe ser exactamente igual a su etiqueta de clase. Esto es difícil de conseguir debido a la variación inherente entre los individuos de la misma clase. El Q^2 depende de la separación entre las clases pero también de la variabilidad dentro de la clase (Westerhuis *et al.*, 2008).

La calidad de los modelos de clasificación también se pueden medir utilizando la tasa de verdaderos positivos (VP), que denota las observaciones de un grupo que se clasifican correctamente como miembros de ese grupo, la tasa de verdaderos negativos (VN), que denota las observaciones que no son miembros de un grupo específico y que están correctamente clasificados como tales. El resto de clasificaciones incorrectas se definen ya sea como falsos positivos (FP) o falsos negativos (FN). Cuando todas las muestras se han predicho y asignado a una clase, el número total de VP , VN , FP y FN se pueden calcular para crear una matriz de confusión que resume la capacidad de predicción del modelo (Broadhurst & Kell, 2006). Varios criterios pueden ser derivados de la matriz de confusión, la sensibilidad (S) y especificidad (Sp) (Altman & Bland, 1994) son los más utilizados. La sensibilidad es una medida de qué tan bien el modelo es capaz de clasificar correctamente las muestras positivas en la clase de los positivos (Ecuación 1.8), mientras que la especificidad mide que tan bien el modelo puede predecir muestras negativas en la clase de negativos (Ecuación 1.9).

$$S = \frac{VP}{(VP+FN)} \quad Ec. 1.8$$

$$S_p = \frac{VN}{(FP+VN)} \quad \text{Ec. 1.9}$$

Debido a que los anteriores parámetros por sí solos no son suficientes para evaluar la calidad y fiabilidad del modelo obtenido, según lo señalado por Westerhuis y colaboradores (2008), la prueba de permutación es una estrategia necesaria que puede evaluar si la clasificación específica de los individuos de los dos grupos es significativamente mejor que cualquier otra clasificación al azar en dos grupos arbitrarios. En una prueba de permutación, se calcula un nuevo modelo de clasificación después de permutar las etiquetas de las clases, osea las variables de clasificación se asignan aleatoriamente entre las distintas muestras del modelo, siendo el resultado esperado que la capacidad predictiva de los modelos aleatorios sea significativamente inferior a la obtenida para el modelo original.

Los parámetros de calidad obtenidos con y sin las etiquetas permutadas se comparan mediante la evaluación del *p-value* (Ecuación 1.10) definido como:

$$p(I) = \frac{1 + \sum_{i=1}^N (I_{p,i} \geq I_{np})}{N} \quad \text{Ec. 1.10}$$

donde, $I_{p,i}$ es el valor del indicador I obtenido para cada permutación aleatoria de las etiquetas categóricas y I_{np} es el valor del indicador obtenido sin permutación. Un modelo es estadísticamente significativo cuando el *p-value* es inferior al nivel de significación (0.05). Finalmente, con el modelo optimizado y validado se puede emplear para estimar las clases de nuevas muestras a partir de las señales analíticas obtenidas previamente.

2. METODOLOGÍA

2.1 RECOLECCIÓN DE MUESTRAS

2.1.1 MODELO DE CLASIFICACIÓN DEL ORIGEN BOTÁNICO Y GEOGRÁFICO DE CAFÉ COLOMBIANO

Las muestras de café finamente molidas (tamaño de partícula 0.075 mm - malla No. 200) fueron proporcionadas por ALMACAFÉ S.A., Bogotá (Colombia) en dos presentaciones (café verde y tostado molido) de las especies *Coffea arabica* y *Coffea canephora* var. *robusta*, y diferentes orígenes geográficos (Asia, África, América). Las 340 muestras usadas para construir el modelo de clasificación, recolectadas entre los años 2011 a 2013, representan a 24 países, tal como se ilustra en la Figura 2.1. La descripción de cada una de ellas se encuentra en el Anexo 6.1.

Las muestras de café tostado se obtuvieron con un grado de tostado medio, lo que implica una temperatura de 200 °C durante 7 min. Se emplearon muestras de 7 regiones cafeteras de Colombia localizadas en las laderas de las tres cordilleras que atraviesan el país de sur a norte (Nariño, Cauca, Huila, Tolima, Caldas, Santander, Cundinamarca y Magdalena), que representan cultivos sembrados con *C. arabica* que se encuentran entre los 1.200 y 1.800 msnm, con temperaturas templadas que oscilan entre los 17 y 23 °C, con precipitaciones cercanas a los 2.000 mm anuales y con un proceso de transformación del fruto por vía “húmeda”.

Café Arábigo Colombiano y Café Robusta seleccionados al azar fueron mezclados para obtener 95 muestras con contenido de café Robusta entre 5 a 90 %. Además, ALMACAFÉ entregó 35 muestras de café tostado catalogadas como sospechosas de adulteración para ser evaluadas con el modelo.

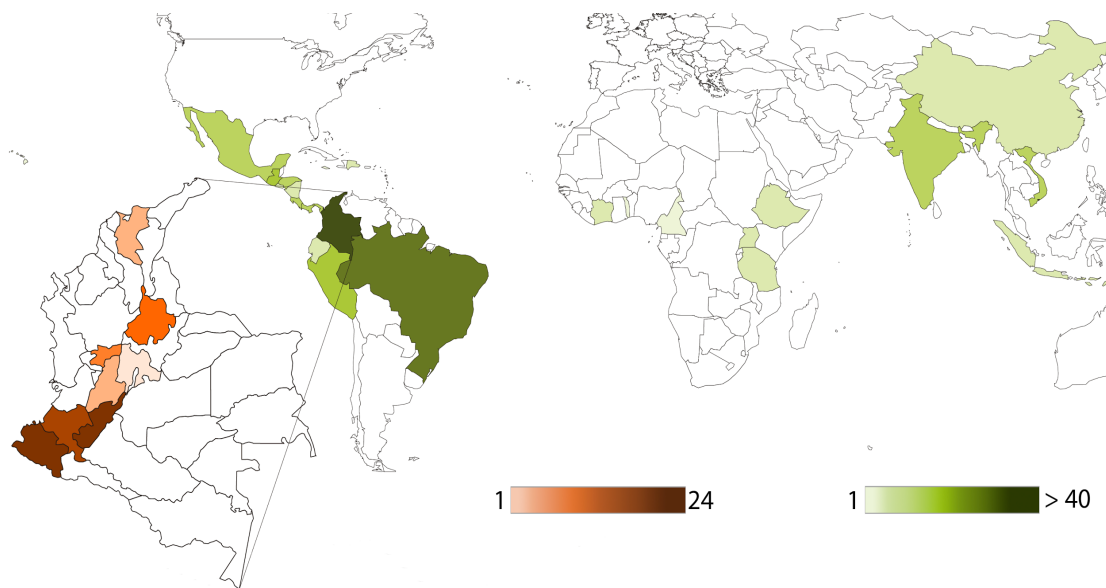


Figura 2.1. Distribución geográfica del origen de las muestras utilizadas para crear un modelo de café Colombiano. Los datos usados para crear esta figura se encuentran en las tablas del Anexo 6.1.

2.1.2 EVALUACIÓN DEL EFECTO DEL GRADO DE TOSTADO SOBRE EL MODELO DE CLASIFICACIÓN

Las muestras de café verde fueron tostadas en un tostador giratorio de laboratorio con circulación de aire caliente (mod. BRZ 4, Probat, Emmerich, DE). Las muestras usadas para evaluar el efecto del grado de tostado en el modelo construido se tostaron durante tiempos diferentes (9 a 13 min) a una temperatura de 200 °C para obtener los grados de tostado medio (9 min), oscuro medio (10 min), oscuro (11 min), y muy oscuro (13 min), y para evaluar la temperatura de tostado se tostaron de 11 a 16 min a una temperatura de 180 °C para obtener los grados de tostado medio (11 min), oscuro medio (13 min), oscuro (14 min), y muy oscuro (16 min). El grado de tostado del café se determinó con el Sistema de Clasificación de Color de Café Tostado desarrollado por la Asociación Americana de Cafés Especiales (SCAA, por su sigla en inglés, Specialty Coffee Association of America) (Agtron-SCAA, Reno, NV, 1995) (Staub, 1995), el cual consiste en clasificar el grado (color) de diferentes tipos de café tostado a partir de la comparación de 8 discos de color numerados (color de referencia) con la muestra de café finamente molida y tostada, de esta forma se asigna al café tostado un

número aproximado en la llamada escala Gourmet Agtron que va desde el # 95 (tostado muy claro) en intervalos de 10 hasta el # 25 (tostado muy oscuro). Se obtuvieron grados de tostado mediante esta clasificación de tostado medio (disco # 55, 22.5 L*), tostado oscuro medio (disco # 45, 21.0 L*), tostado oscuro (disco # 35, 17.0 L*) y tostado muy oscuro (disco # 25, 15.5 L*). L* es la coordenada de luminosidad de la escala CIELAB (CIE - Commission Internationale de l'Eclairage- Comisión Internacional de Iluminación) (Figura 2.2).



Figura 2.2. Grado de tostado del café: (A) tostado medio (22.5 L*), (B) tostado oscuro medio (21.0 L*), (C) tostado oscuro (17.0 L*), (D) tostado muy oscuro (15.5 L*)

2.1.3 EVALUACIÓN DE LAS TÉCNICAS ANALÍTICAS GC-MS, GC-C-IRMS Y RMN PARA DETERMINAR EL ORIGEN GEOGRÁFICO DEL CAFÉ

El subconjunto de 34 muestras de café tostado (*Coffea arabica* L.) de 3 países de Sur América (Colombia, Brasil y Perú) colectadas durante un periodo de 2 años, 2012 a 2013, fue seleccionado al azar para conservar la mayor parte de las fuentes de variación, con el fin de evaluar tres técnicas analíticas. Las muestras entregadas por Almacafé S.A. (Colombia) fueron distribuidas de la siguiente manera: 15 de Colombia, 11 de Brasil y 8 de Perú. Estas muestras fueron analizadas por GC-MS y GC-C-IRMS en el Instituto de Ciencias Forenses de la Escuela de Ciencias Criminalísticas, Universidad de Lausanne, Suiza. Estas mismas muestras también fueron analizadas por RMN 400 MHz en el Departamento de Química, Universidad del Valle, Colombia.

2.2 DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR RMN

Se realizaron pruebas con diferentes solventes para extraer los metabolitos del café. Los espectros obtenidos se compararon y evaluaron sobre la base de dos criterios: costo y simplicidad, y la cantidad de señales presentes. Se pesó 100 mg de café verde molido y se adicionó 1.0 mL de solvente deuterado, se agitó y filtró. Los solventes deuterados usados para las pruebas fueron: óxido de deuterio (D_2O), cloroformo ($CDCl_3$), dimetilsulfóxido (DMSO) y metanol (CH_3OD). Para establecer el tiempo y temperatura de extracción de los metabolitos del café se realizaron ensayos, con tiempos entre 2 y 10 minutos, y con calentamiento a 90 °C y sin calentamiento. Además, se evaluaron dos procesos de purificación de los extractos obtenidos, debido a que las partículas suspendidas en el extracto afectan la línea base y la reproducibilidad de los espectros, por consiguiente afectan el resultado del análisis estadístico de los datos, el primero se realizó con filtros de membrana de 0.45 μm y el segundo con centrifugación, los ensayos se realizaron varias veces para la misma muestra y en diferentes días.

Para el caso de utilizar agua como solvente de extracción, se debe controlar el pH de la muestra para que no influya sobre la posición de las señales en particular la señal del agua ya que conduciría a degradar la calidad de la supresión de la señal del agua en el análisis por RMN, esto se debe a que el pH de la muestra es un factor crítico para la estabilidad de las señales debido a los grupos ácidos o básicos de los metabolitos típicos (Tynkkynen *et al.*, 2009). El uso de una solución tampón puede reducir esta variación y garantizar una mejor reproducibilidad. Se realizaron ensayos con solución tampón de fosfato a pH 5 y 7 a una concentración de 100 mM. Se debe asegurar que las soluciones tampón preparadas sean siempre las mismas para no generar variación en la posición de las señales, por lo tanto se realizó un proceso de

estandarización de la solución tampón por ^{31}P -RMN, con el objeto de asegurar siempre el mismo valor de pH.

Los protocolos obtenidos para la preparación de las muestras de café tanto en solución tampón como en metanol sin sustitución isotópica de los protones con deuterio se encuentran en los Anexos 6.2 y 6.3, respectivamente.

2.3 DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR GC-MS

Se desarrolló y optimizó las condiciones experimentales para obtener el extracto metabólico de las muestras de café tostado para el análisis por GC-MS. Se realizaron pruebas variando las condiciones experimentales, tales como solvente de extracción (cloroformo, diclorometano, hexano, metanol y acetato de etilo), tamaño de muestra, tipo de extracción (vortex o ultrasonido), tiempo de extracción y método de purificación (filtración o centrifugación). Los criterios para seleccionar el método de extracción fueron: selección del solvente por el mayor número de metabolitos extraídos, menor tiempo de preparación y mejor resolución en el perfil metabolómico, en el Anexo 6.4 se presenta el protocolo para la preparación de muestras.

2.4 DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE CAFEÍNA PARA EL ANÁLISIS DE $\delta^{13}\text{C}$ POR CG-C-IRMS

Se desarrolló y optimizó las condiciones experimentales para obtener el extracto metabólico de las muestras de café tostado para el análisis por GC-C-IRMS en base a la metodología reportada por Weckerle y colaboradores (2002). Se empleó un método de extracción líquido-líquido y se realizaron pruebas con diferentes solventes de

extracción y tamaños de muestras. En el Anexo 6.5 se presenta el mejor método de extracción de cafeína para café tostado obtenido para el análisis por GC-C-IRMS.

2.5 ANÁLISIS RMN

Los experimentos de RMN se llevaron a cabo midiendo la señal del protón con un espectrómetro Bruker Avance II 400 MHz (Bruker BioSpin, Rheinstetten, Alemania) operado a una frecuencia de 400.02 MHz, equipado con una sonda directa con gradientes de triple eje, accesorios de tuning y matching automático, BVT-1000 y BCU I -40/50 para regulación de temperatura, y un automuestreador SampleXpress Lite (Bruker BioSpin, Rheinstetten, Alemania). Los datos espectrales se guardaron con el software Bruker TopSpin (v 2.1, Bruker BioSpin, Rheinstetten, Alemania). Los extractos se analizaron a 300 K en modo totalmente automático. Después se insertó cada tubo con el extracto al imán con un largo retraso “*delay*” (180 s) antes de la adquisición para asegurar que la temperatura del extracto en el interior del imán se haya equilibrado. Tuning, matching (*ATMA*), locking (*LOCK*), shimming (*TOPSHIM*), calibración del pulso de 90° (*PULSECAL*) y pulso de saturación (25 Hz) se lograron utilizando rutinas estándar de Bruker.

Antes de realizar la adquisición de cada conjunto de experimentos, se debió optimizar y calibrar todos los parámetros críticos en este tipo de estudios metabolómicos, como temperatura, shims, frecuencia de la señal del agua (O1), etc. De manera general, los protocolos de optimización comprenden los siguientes pasos:

a. Calibración de la temperatura con una muestra de referencia de metanol deuterado (99.8 %). Secuencia de pulso: zg30

Para establecer la temperatura de trabajo se utilizó una muestra de referencia de metanol deuterado que permitió seguir las fluctuaciones de la temperatura durante varios días de experimentos. Se usó la calibración de temperatura de Amman y colaboradores (1982) (Ecuación 2.2).

$$T [K] = -21.85\Delta^2 - 36.54\Delta + 409.0 \quad Ec. 2.2$$

Donde, T = Temperatura, en K. Δ = Diferencia de frecuencia entre las señales -CH₃ y -OH del metanol.

b. Optimización de los shims y de la frecuencia de la señal del agua (O1) con una muestra patrón de sacarosa (2 mM sacarosa en 90 % H₂O + 10 % D₂O). Secuencia de pulso: zgpr

c. Optimización de los shims y de la frecuencia de la señal del agua (O1) para una muestra representativa del estudio (muestra preparada con estándares de cafeína, ácidos clorogénicos, sacarosa y trigonelina). Secuencia de pulso: zgpr.

d. Comprobación del ajuste de los parámetros en un espectro 1D-NOESY de la muestra representativa del estudio y evaluación de las siguientes características:

- Línea base plana
- Buena corrección de la fase
- Ajuste de los shims: Señal simétrica del TMS, con señales satélite pronunciadas, y una anchura de señal inferior a 1 Hz.
- Buena supresión de la señal del agua: La señal residual del agua debe ser significativamente menor que la obtenida en el correspondiente espectro zgpr, con una anchura de línea inferior a 100 Hz.

Para la adquisición de los espectros de los extractos de café obtenidos con solución tampón se utilizó el experimento 1D-NOESY-Presat, secuencia de pulso: noesygppr1d.comp. En la Figura 2.3 se presenta la región de la supresión de la señal del agua en un extracto de café verde.

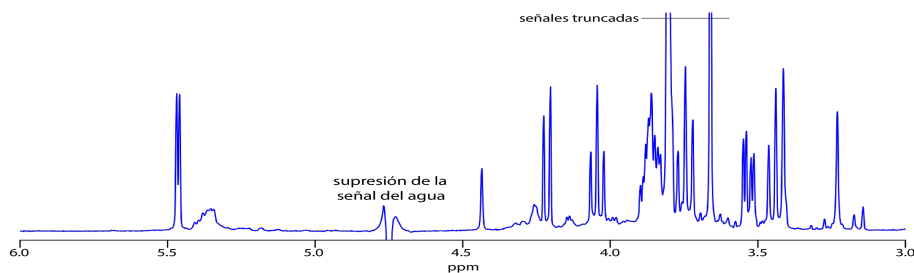


Figura 2.3. Región de la supresión de la señal del agua en un extracto de café verde obtenido con solución tampón.

Los espectros de los extractos metanolicos de café se obtuvieron utilizando la misma secuencia 1D-NOESY-Presat arriba mencionada pero con supresión múltiple, tanto de las señales de metanol como del agua, Figura 2.4. Se utilizó una secuencia de experimentos ^1H -RMN, iniciando con el experimento *zg30* para medir la frecuencia de las señales de metanol y agua, luego, usando las frecuencias medidas, se realizó el segundo experimento *zgpsd0* para realizar la presaturación y se volvió a medir las frecuencias de las señales de metanol y agua, por último el experimento *noesygpps1d* utilizó las frecuencias medidas en el segundo experimento y se realizó la supresión multiple definitiva de las señales de metanol y agua. Esta secuencia de experimentos de RMN se realizó de manera automática y permitió la visualización de las señales de los metabolitos en las muestras de café. Los parámetros de los tres experimentos 1D son presentados en el Anexo 6.6.

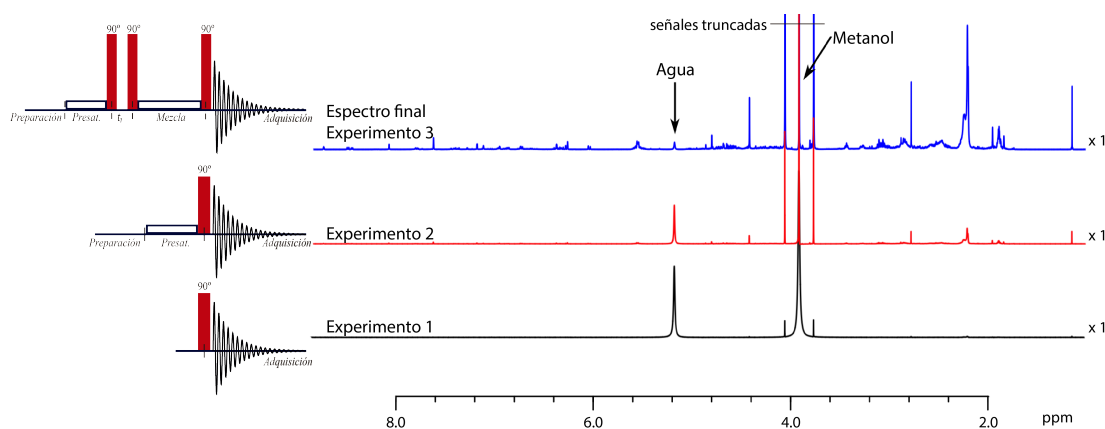


Figura 2.4. Región de la supresión de las señales de metanol y agua en un extracto de café tostado. Experimento 1: *zg30*, Experimento 2: *zgpsd0*, Experimento 3: *noesygpps1d*.

2.6 ANÁLISIS GC-MS

El análisis por GC-MS fue llevado a cabo en un cromatógrafo de gases Agilent Technologies 6809N Network (Agilent Technologies, Palo Alto, CA) acoplado a un detector selectivo de masas HP 5975 y un automuestreador HP 7673 (Agilent Technologies, Palo Alto, CA). Las condiciones de operación son presentadas en el Anexo 6.7. La adquisición y evaluación de los datos GC-MS fueron realizados con el

software ChemStation (Agilent Technologies, Palo Alto, CA).

Los metabolitos obtenidos en el cromatograma fueron identificados comparando sus espectros de masas con algunos de los estándares puros y la librería Wiley (McLafferty & Staufer, 1989). Los compuestos identificados se cuantificaron por integración de sus señales. Cada integral se normalizó con respecto a la suma de la integral presente en cada traza, por lo tanto se calcularon las intensidades relativas para cada compuesto (Zambonin et al., 2005). Cada muestra de café tostado fue analizada por duplicado.

2.7 ANÁLISIS GC-C-IRMS

Las medidas de los valores de $\delta^{13}C$ de la cafeína fueron llevados a cabo en un espectrómetro de masas de relación isotópica Delta V Advantage Isotope Ratio Mass Spectrometer (IRMS) system (ThermoFisher Scientific, Bremen, Alemania) acoplado a un cromatógrafo de gases Trace GC Ultra Gas Chromatograph via a GC-C/TC III interface (ThermoFisher Scientific, Bremen, Alemania), operado en el modo “Combustión” (C). Las muestras fueron inyectadas con automuestreador TriPlus™ (ThermoFisher Scientific, Bremen, Alemania). Las condiciones de operación son presentadas en el Anexo 6.8.

Los compuestos orgánicos de la muestra que eluyeron de la columna cromatográfica fueron oxidados a gas CO₂ en un reactor de combustión. El eluyente fue purificado a través de una trampa de agua y transferido a la fuente de iones del espectrómetro de masas para determinar la relación isotópica de carbono mediante la medición de las masas 44, 45 y 46. Los valores $\delta^{13}C$ de las muestras se calcularon en relación a la medida de un gas de referencia (CO₂) de composición isotópica conocida, que se introdujo con pulsos de 20 s de duración a 6 tiempos diferentes durante el curso de la separación cromatográfica. La oxidación del reactor de combustión se realizó durante 4 h después de cada lote de 20 análisis.

El símbolo δ es la notación estándar para expresar la relación isotópica de carbono. Se define como partes por mil de la desviación de la composición isotópica de Viena Pee Dee Belemnite, VPDB (ecuación 2.2) (Craig, 1957) y se calculó con el software ISODAT 2.5 (ThermoFisher Scientific, Bremen, Alemania).

$$\delta^{13}C_{\text{muestra}} (\text{‰}) = [(R_{mu} / R_{st}) - 1] * 10^3 \quad \text{Ec. 2.2}$$

Donde, R_{mu} y R_{st} corresponde a la relación isotópica ($^{13}C/^{12}C$) de la muestra y el estándar interno (st): ácido tetradecanoico metil éster con un valor de relación isotópica $\delta^{13}C = -29.98$, respectivamente. Cada muestra de café tostado fue analizada por duplicado.

2.8 PROCESAMIENTO DE LOS DATOS OBTENIDOS POR RMN

Fue necesario realizar un procesamiento de los datos espectrales antes de realizar el análisis estadístico, esto con el fin de reducir diferencias e influencias externas. Una vez adquiridos los experimentos se realizó automáticamente la transformada de Fourier, corrección de la línea base y corrección de fase, para obtener una representación interpretable de los datos. Luego se realizó una exclusión de las regiones del espectro que no contienen información metabólica y que son dependientes de los parámetros experimentales, las regiones de las señales del metanol (δ 3.14 - 3.55 ppm) y TMS (δ -0.20 – 0.20 ppm), así como las regiones espectrales de los extremos. Para reducir los errores generados por pequeños desplazamientos en la posición de las señales se empleó el método de disminución de la resolución espectral (*binning*), con *bins* o *buckets* (subdivisión e integración del espectro en pequeñas regiones) de anchura igual a δ 0.04 ppm en la región espectral comprendida entre δ 0.5 – 9.6 ppm.

2.9 ANÁLISIS ESTADÍSTICO MULTIVARIADO

2.9.1 ANÁLISIS NO SUPERVISADOS

Se realizó PCA para el conjunto de datos para obtener una visión global de la variabilidad entre las muestras, su correspondiente gráfico de puntuaciones permitió identificar los posibles agrupamientos o tendencias dentro del grupo de muestras y la elipse T^2 de Hotelling permitió identificar las muestras que presentaban un comportamiento atípico en comparación a su grupo y que se encontraban fuera del intervalo de confianza del 95 %. La identificación de estas muestras fue importante porque podrían representar fuentes de heterogeneidad no deseadas, causando interferencia en los resultados degradando la capacidad predictiva del modelo.

2.9.2 ANÁLISIS SUPERVISADOS - MODELOS DE CLASIFICACIÓN

Para la construcción del modelo de clasificación PLS-DA u OPLS-DA, método supervisado de reconocimiento de patrones que proporcionó la mejor descripción de los datos así como la mejor discriminación entre muestras de dos clases diferentes. Se realizó la clasificación binaria entre origen botánico, *C. arabica* contra *C. canephora* var. *robusta* y mezcla de estas especies, de aquí en adelante serán denominadas “No *C. arabica*” y luego para la especie *C. arabica* se clasificó entre origen geográfico, “café Colombiano” y “otros orígenes”. Para evaluar la calidad de la capacidad predictiva del modelo de clasificación se empleó el procedimiento de doble validación cruzada (ver Sección 1.5), se escogió al azar un conjunto de datos que se colocaron a un lado y no fueron utilizados para la construcción del modelo (conjunto de validación externa). El resto del conjunto de datos fueron usados para construir y optimizar el modelo, se escogió al azar un conjunto de entrenamiento (80 %) y otro conjunto de prueba (20 %), de esta forma el conjunto de entrenamiento permitió desarrollar una serie de modelos PLS-DA que fueron usados para predecir sobre el conjunto de prueba. Este procedimiento se repitió 100 veces. Cada modelo independiente PLS-DA obtenido fue utilizado para predecir sobre el conjunto de

validación externa. Varios parámetros fueron usados para estimar la calidad de los modelos de clasificación. De la matriz de confusión se derivaron parámetros como sensibilidad (S) y especificidad (Sp), donde S es una medida de qué tan bien el modelo es capaz de clasificar correctamente las muestras de la clases “*C. arabica*” o “*café Colombiano*”, y Sp es una medida de lo bien que el modelo puede predecir muestras de la clases “No *C. arabica*” u “*otros orígenes*”. Valores R^2 y Q^2 , donde R^2 es la proporción de la varianza en los datos explicada por el modelo, e indica la bondad del ajuste y Q^2 es la proporción de la varianza en los datos predecibles por el modelo y refleja la exactitud de la predicción (Szymańska *et al.*, 2012). El resultado final de estos parámetros se obtuvieron a partir de la media aritmética de los 100 valores obtenidos. Además, se realizó el mismo procedimiento para obtener el p -value de 100 modelos con etiquetas de clases permutadas, los cuales también fueron reportados el valor medio aritmético. Los resultados se visualizaron mediante el gráfico de puntuaciones que mostró la variabilidad de las muestras y con el gráfico de cargas que representó las variables que maximizan las diferencias entre las clases. Además, otra forma de facilitar la interpretación de los resultados fue mediante el gráfico de cargas con escala de colores, que presentó el peso discriminante de cada variable de acuerdo a la escala de colores ubicada al lado derecho de la gráfica, por ejemplo, el color rojo indicaba contribución significativa a la separación entre las dos clases que el color azul (Know *et al.*, 2015). Finalmente, las muestras sospechosas de fraudes fueron analizadas cada una con los 100 modelos de clasificación obtenidos y se reportó la frecuencia con que cada modelo las clasificó correctamente. Estos análisis se realizaron utilizando códigos de programación realizados en el Grupo de Investigación para el software R (R Development Core Team, 2013) y basados en el paquete estadístico *caret* (Max, 2013).

2.9.3 ANÁLISIS SUPERVISADOS - MODELOS DE CUANTIFICACIÓN

El análisis de regresión PLS se realizó con el fin de obtener un modelo cuantitativo para la predicción del contenido de café Robusta cuando es mezclado con café

Arábica basándose en la información espectral del café tostado. También se realizó un análisis de regresión PLS para obtener un modelo cuantitativo para la predicción del grado de tostado del café, en unidades de L^* (luminosidad). PLS es un método basado en variables latentes, particularmente útil cuando se trata de datos ruidosos y colineales (Rannar, Lindgren, Geladi, & Wold, 1994). Las variables latentes son direcciones ortogonales en el espacio de los predictores caracterizados por la covarianza máxima con la variable de respuesta seleccionada. El número óptimo de variables latentes que proporciona el mínimo error, evaluado como el error cuadrático medio de la validación cruzada (*RMSECV*), fue seleccionado. Estos análisis se realizaron utilizando códigos de programación realizados en el Grupo de Investigación para el software R (R Development Core Team, 2013) y basados en el paquete estadístico *pls* (Mevik, Wehrens, & Liland, 2013).

3. RESULTADOS Y DISCUSIÓN

3.1 DETERMINACIÓN DEL PROTOCOLO DE EXTRACCIÓN DE METABOLITOS DE CAFÉ PARA EL ANÁLISIS POR RMN

Definir el protocolo de extracción de las muestras de café verde y tostado es muy importante para los estudios metabolómicos, debido a que se debía asegurar la reproducibilidad de los espectros obtenidos. Por el alto costo del solvente deuterado y el gran número de muestras para analizar se realizó la extracción de los metabolitos en solvente sin sustitución isotópica y la adquisición de los espectros totalmente automática. En primer lugar, fue importante seleccionar el mejor solvente de extracción debido a que no se conocía de antemano que tipo de moléculas serían las de mayor peso para discriminar origen botánico y geográfico. En la Figura 3.1 se observa la diferencia en la cantidad de señales a lo largo del desplazamiento químico cuando son superpuestos los espectros de los extractos de café verde obtenidos con los solventes cloroformo- d_1 (azul), dimetilsulfóxido- d_6 (rojo), metanol- d_4 (verde) y óxido de deuterio (negro).

Se seleccionaron los solventes óxido de deuterio y metanol- d_4 para obtener más experimentos, ya que estos solventes presentaron mayor cantidad de señales entre 6 y 9 ppm. Esta región se caracteriza por la presencia de metabolitos como cafeína, trigonelina y ácidos clorogénicos (Bosco, Toffanin, de Palo, Zatti, & Segre, 1999; Wei, Furihata, Hu, Miyakawa, & Tanokura, 2010; Wei, Furihata, Hu, Miyakawa, & Tanokura, 2011), los cuales en diversos estudios han demostrado ser variables importantes y pueden contribuir a la discriminación (Andrade *et al.*, 1998; Martín *et*

al., 1998; Casal *et al.*, 2000; Tavares & Ferreira, 2006; Mendonça *et al.*, 2008; Alonso-Salces *et al.*, 2009).

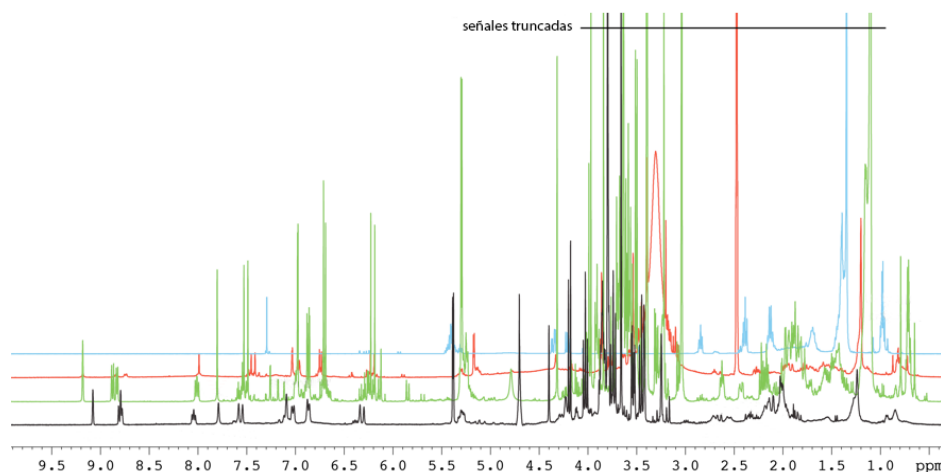


Figura 3.1. Espectros de extractos de café verde obtenidos con los solventes óxido de deuterio (negro), metanol- d_4 (verde), dimetilsulfóxido- d_6 (rojo) y cloroformo- d_1 (azul).

La extracción de los metabolitos del café se realizó con solventes sin sustitución isotópica, debido al alto costo de los solventes deuterados. Además, se realizó la supresión de la señal del solvente en el análisis por RMN. Inicialmente se seleccionó la solución acuosa para realizar la extracción por presentar el mayor número de señales y ser el solvente más reportado para el análisis de café (Charlton, Farrington, & Brereton, 2002; Wei *et al.*, 2010; Consonni, Cagliani, & Cogliati, 2012), se estableció el tiempo de extracción de 2 minutos y sin calentamiento, debido a que se encontró que al aumentar el tiempo y la temperatura no afectaban en la intensidad de las señales de las regiones de interés del espectro, dando como ventaja tiempo de análisis corto y menos variables para controlar, como la temperatura.

Además de los anteriores parámetros, para asegurar la reproducibilidad de las señales espectrales de los extractos, se seleccionó el mejor método de purificación de los extractos, ya que la calidad de la línea base espectral depende fuertemente de la concentración de las partículas en suspensión, esto puede ser explicado por el efecto de las partículas pequeñas en la homogeneidad del campo y por lo tanto, sobre la

eficiencia del campo de pre-saturación. Por lo tanto, para las muestras de café verde fue necesario realizar doble centrifugación para eliminar las partículas en suspensión y así asegurar una buena reproducibilidad en las señales del espectro, mientras que para las muestras de café tostado con una sola centrifugación fue suficiente, esto coincide con la mayoría de las metodologías reportadas (Wei, Furihata, Hu, Miyakawa, & Tanokura, 2010; Wei, Furihata, Hu, Miyakawa, & Tanokura, 2011).

En cuanto a la selección de la solución tampón, se obtuvo mejores resultados con la solución tampón de fosfato de pH 7.0 (100 mM) ya que permitió una mayor eficiencia en la supresión de la señal del agua debido a que no presentó cambios en el desplazamiento químico de las señales, ni atenuación de las mismas y además por ser el más empleado en estudios de café (Ciampa, Renzi, Taglienti, Sequi, & Valentini, 2010; Consonni, Cagliani, & Cogliati, 2012; Cagliani, Pellegrino, Giugno, & Consonni, 2013).

Se debe tener en cuenta que cuando se trabaja con solución tampón, cada solución nueva que se prepare debe tener, además del mismo valor de pH a la solución anterior (pH 7.0, 100 mM) la misma composición, por lo tanto no era suficiente ajustar el pH con el pHmetro sino que cada nueva solución debía ser siempre estandarizada por ^{31}P -RMN con respecto a la solución anterior para garantizar la reproducibilidad de la señal de las soluciones tampones preparadas. El proceso de estandarización consistió en adquirir los espectros por ^{31}P -RMN de las soluciones iniciales de HPO_4^{2-} y H_2PO_4^- , así como de las soluciones tampón nueva y la preparada anteriormente, la señal de la solución tampón nueva debía coincidir en la misma región del espectro de la solución tampón preparada anteriormente, en caso contrario se debió ajustar la señal adicionando a la nueva solución tampón volúmenes pequeños de alguna de las dos soluciones iniciales con la cual fue preparada dependiendo hacia donde se deseaba desplazar la señal (Figura 3.2).

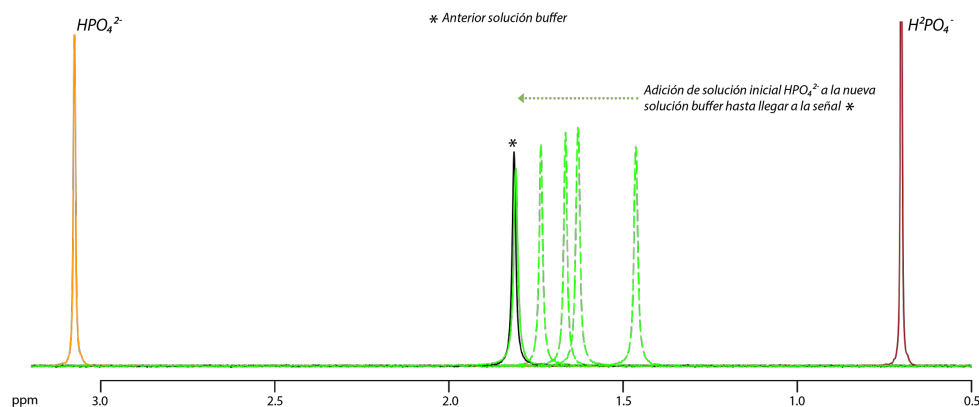


Figura 3.2. Estandarización por ^{31}P -RMN de una nueva solución tampón de fosfato (verde) preparada con las soluciones iniciales de HPO_4^{2-} (naranja) y H_2PO_4^- (marrón), donde se observa el desplazamiento de la señal de la nueva solución tampón (verde) hasta ajustarla con la señal de la anterior solución tampón (negro) .

Después de obtener el protocolo de extracción con solución tampón acuoso para muestras de café verde y tostado, se procedió a la automatización para la adquisición de los espectros con rutinas estándar de Bruker. Con los extractos de café verde se observó una rápida degradación (< 1 h) siendo difícil la completa automatización del análisis, ya que la corrida de la secuencia de 16 tubos en el automuestreador tenía un tiempo de duración de 3 h, aproximadamente. El sistema de automuestreador no contaba con un mecanismo de enfriamiento que permitiera mantener estable el extracto durante este tiempo. Por esta razón, al igual que con la solución tampón acuosa, se desarrolló y evaluó un protocolo experimental, rápido y económico, para la extracción de las muestras de café verde y tostado, empleando metanol grado analítico sin sustitución isotópica de los protones de deuterio.

En cuanto a la calibración de la temperatura de trabajo del RMN se obtuvo una temperatura promedio de 300.25 K con una desviación estándar de 0.18. Se determinó que a partir de 300 K, el sistema era capaz de conservar una temperatura estable durante más de 24 horas. Por debajo, el aire (a temperatura ambiente) que

sirve para enfriar no tenía un gradiente de temperatura suficiente para que el sistema se mantuviera estable.

3.2 HUELLA METABOLÓMICA ^1H -RMN DE CAFÉ VERDE Y TOSTADO

Muchas familias de compuestos se lograron observar en los espectros ^1H -RMN de café verde obtenidos tanto con solución tampón acuosa (Figura 3.3) como con metanol (Figura 3.4), los cuales exhiben un patrón complicado de señales. Las resonancias de los protones de los grupos alifáticos se observaron con un considerable solapamiento en la región de campo alto entre 1.7 a 3.0 ppm. Las regiones entre 3.0 y 4.5 ppm contenían señales de los anillos alifáticos de ACQ y sacarosa. Las señales en la región de campo bajo, por debajo de 7.0 ppm se debieron a los protones de los anillos aromáticos y dobles enlaces de ACQ, cafeína y trigonelina. Las señales de la cafeína fueron fácilmente asignadas, siendo muy intensas en el espectro.

En el espectro ^1H -RMN del extracto de café verde obtenido con solución tampón acuoso (Figura 3.3) se asignaron señales de protones en la región de campo alto para grupos metilos o metileno de ácido quínico, ácido málico y ácido cítrico. Algunas señales singlete en 3.23, 3.41 y 4.2 ppm fueron asignadas a los tres grupos N-metilo de la cafeína y la señal en 4.43 al grupo N-metilo de la trigonelina. Las señales a 8.24, 9.04 y 9.33 ppm se obtuvieron a partir del anillo de piridina de trigonelina, la señal en 7.98 ppm se derivó del protón aromático de cafeína; dos grupos de dobletes de centraron alrededor de 6.45 ppm y 7.79 ppm, respectivamente, este patrón es típico de protones olefinicos en una configuración trans de los isómeros de ACQ; y las señales en 7.27 ppm se derivaron de los protones aromáticos de ACQ. Entre 0.60 - 1.30 ppm se evidenciaron amplias señales relativas a los grupos metilo y metileno de los ácidos grasos y triglicéridos, lo que confirmó la presencia de una fracción de lípidos. Estas señales fueron asignadas con base a lo reportado por Wei y colaboradores (2010).

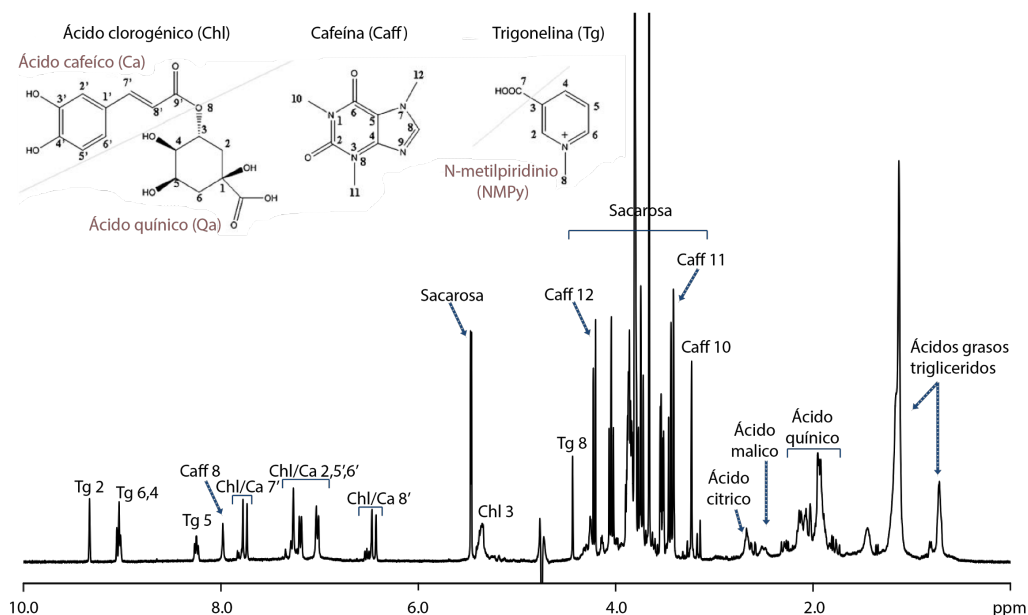


Figura 3.3. Espectro ^1H -RMN de una muestra de café verde obtenido con solución tampón acuosa y asignación de las principales moléculas.

En el espectro ^1H -RMN del extracto de café verde obtenido con metanol (Figura 3.4), se asignaron las señales en 3.96 y 4.43 ppm al grupo N-metilo de la cafeína y trigonelina, respectivamente, las señales del anillo de piridina de la trigonelina en 8.05, 8.88 y 9.20 ppm, la señal del protón aromático de la cafeína en 7.85 ppm. Las señales en 5.34, 6.29, 6.77, 7.04 y 7.56 ppm al grupo de los ácidos clorogénicos y la región entre 0.85 y 1.40 ppm al grupo de los lípidos y triglicéridos. Estas señales fueron asignadas con base a lo reportado por Kwon y colaboradores (2015).

En el espectro de ^1H -RMN de extracto de café tostado obtenido con solución tampón acuosa a pH 7.0 (Figura 3.5) las señales de cafeína, trigonelina, N-metilpiridinio, ácido fórmico, acetato, fracción lipídica y ácidos clorogénicos están presentes y bien resueltas, coincidiendo con lo reportado en la literatura (Bosco, Toffanin, de Palo, Zatti & Segre 1999; D'Amelio, Fontanive, Uggeri, Suggi-Liverani & Navarini, 2009). También se evidenció la presencia de los principales compuestos que componen el café en el espectro ^1H -RMN del extracto metanólico de café tostado

(Figura 3.6).

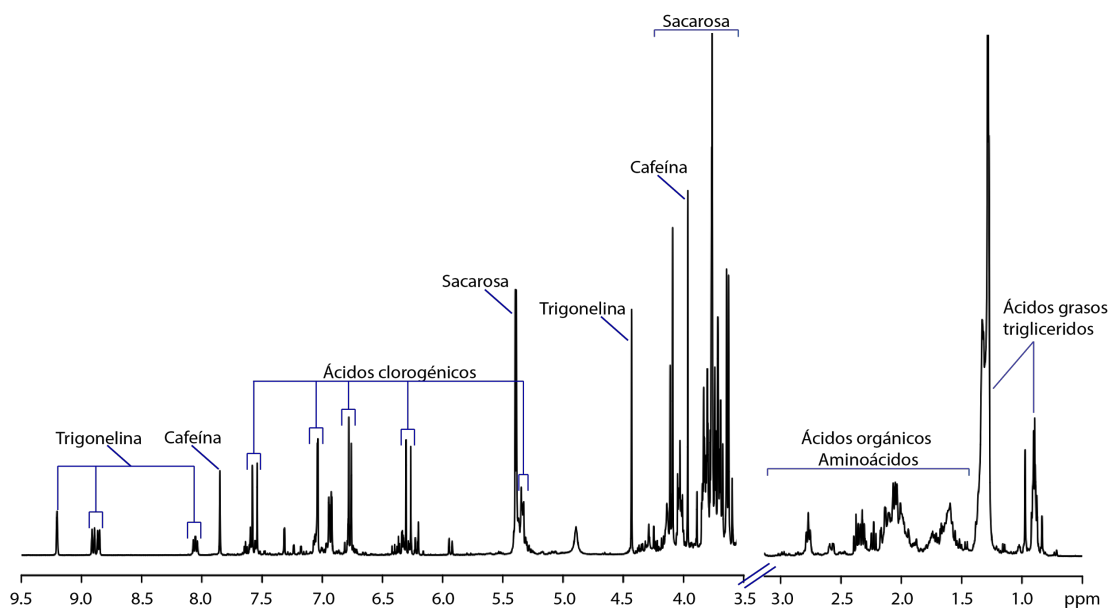


Figura 3.4. Espectro ^1H -RMN de una muestra de café verde obtenido con metanol y asignación de las principales moléculas.

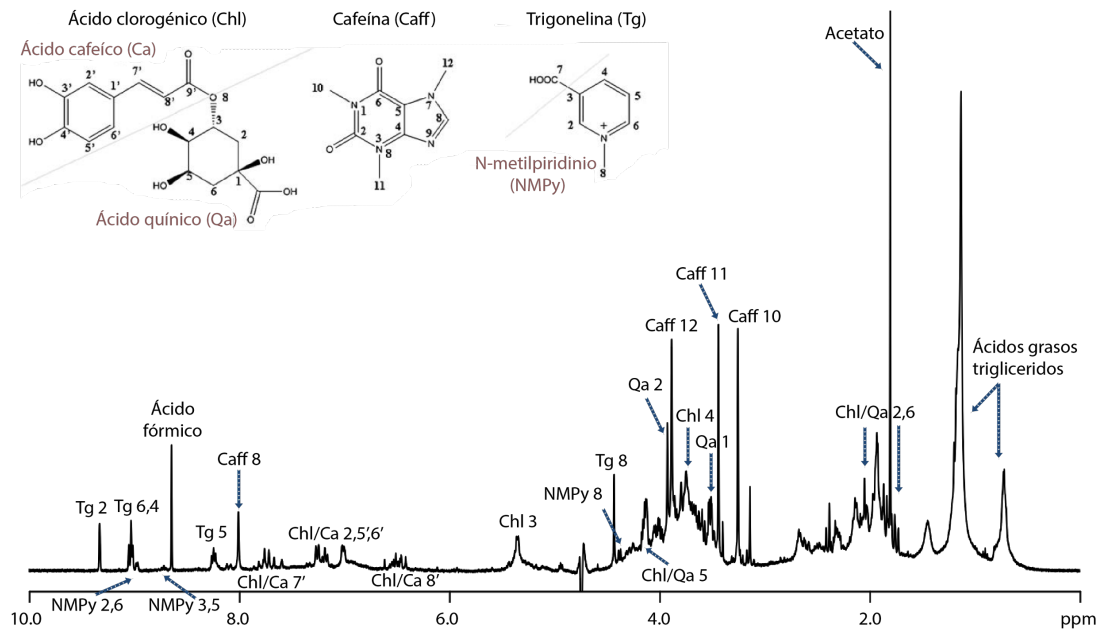


Figura 3.5. Espectro ^1H -RMN de una muestra de café tostado obtenido con solución tampón y la asignación de las principales moléculas.

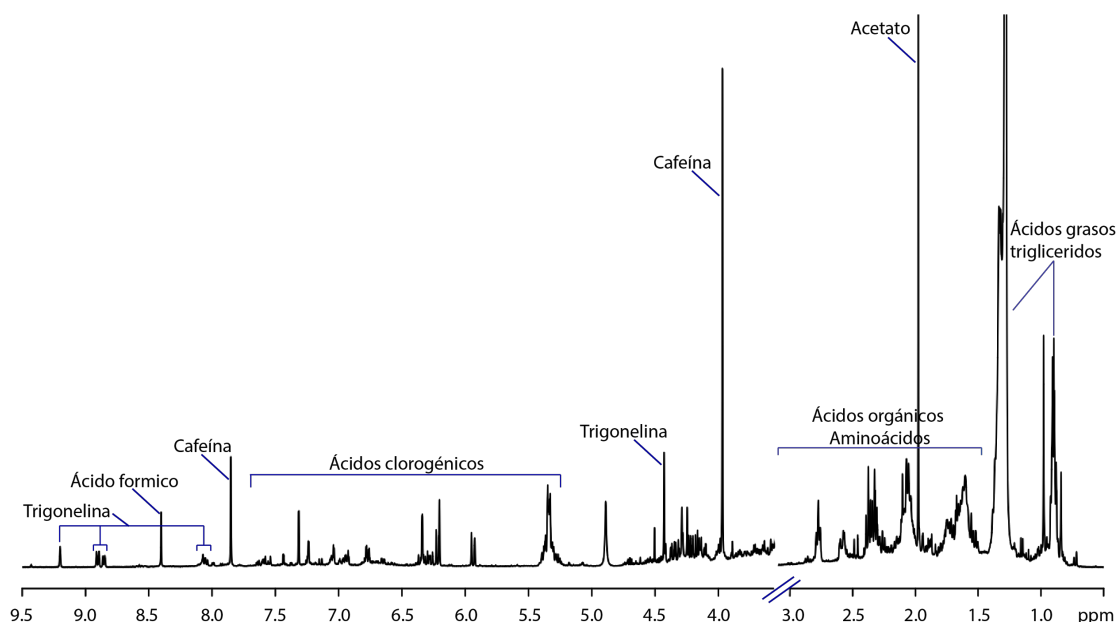


Figura 3.6. Espectro ^1H -RMN de una muestra de café tostado obtenido con metanol y la asignación de las principales moléculas.

Al comparar los perfiles metabolómicos de las especies *C. arabica* y *C. canephora* en extractos metanólicos de café tostado (Figura 3.7) se observaron claras diferencias que han sido evidenciadas en la Tabla 1.2. Una de ellas es la señal a 0.98 ppm presente en muestras de *C. arabica* y ausente en las muestras de *C. canephora*. Las señales en 2.2 - 2.35, 4.1, 3.97 y 7.85 ppm presentaron una mayor intensidad en *C. canephora*, mientras que *C. arabica* las presentaron en 1.5 - 1.8, 2.10, 4.2 - 4.3, 5.9 - 6.4, 7.2 - 7.4 ppm.

Se puede también observar en el perfil metabolómico la presencia de la especie *C. canephora* var. *robusta* cuando es mezclada en una muestra de *C. arabica* (Figura 3.8), algunas regiones del espectro ^1H -RMN evidenciaron cambios en las señales por este tipo de adulteración, por ejemplo las señales de la cafeína ubicadas en 3.96 y 7.85 ppm se vieron afectadas tanto en intensidad como en desplazamiento químico cuando aumenta la proporción de café Robusta en la muestra de café Arábica.

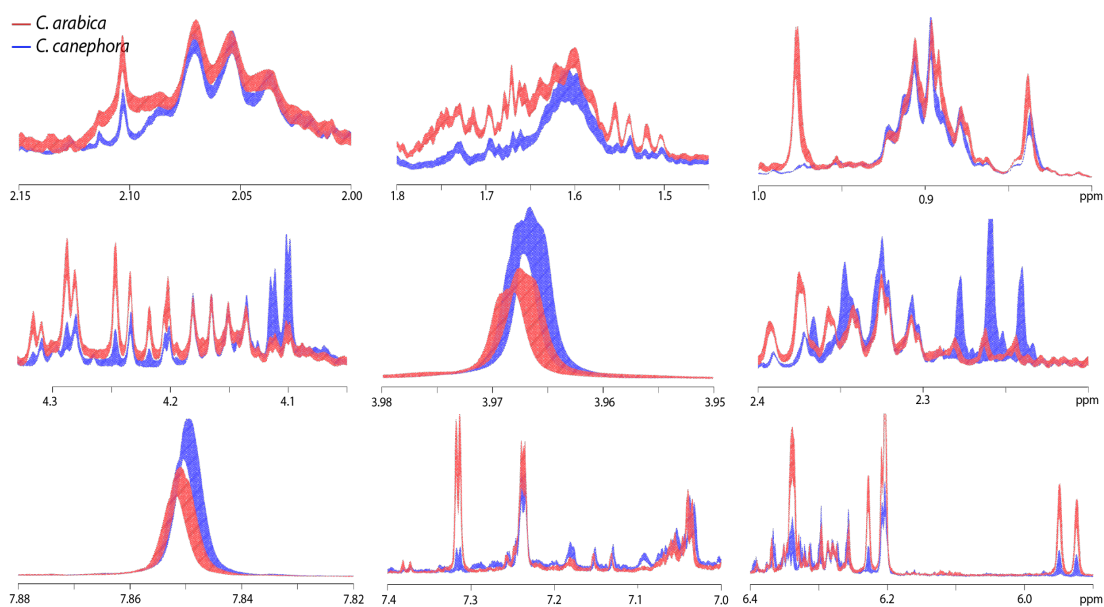


Figura 3.7. Perfil metabolómico ^1H -RMN de las especies *C. arabica* (rojo) y *C. canephora* (azul) en extractos metanólicos de semillas tostadas. En la especie *C. canephora* no hay presencia de la señal en 0.98 ppm.

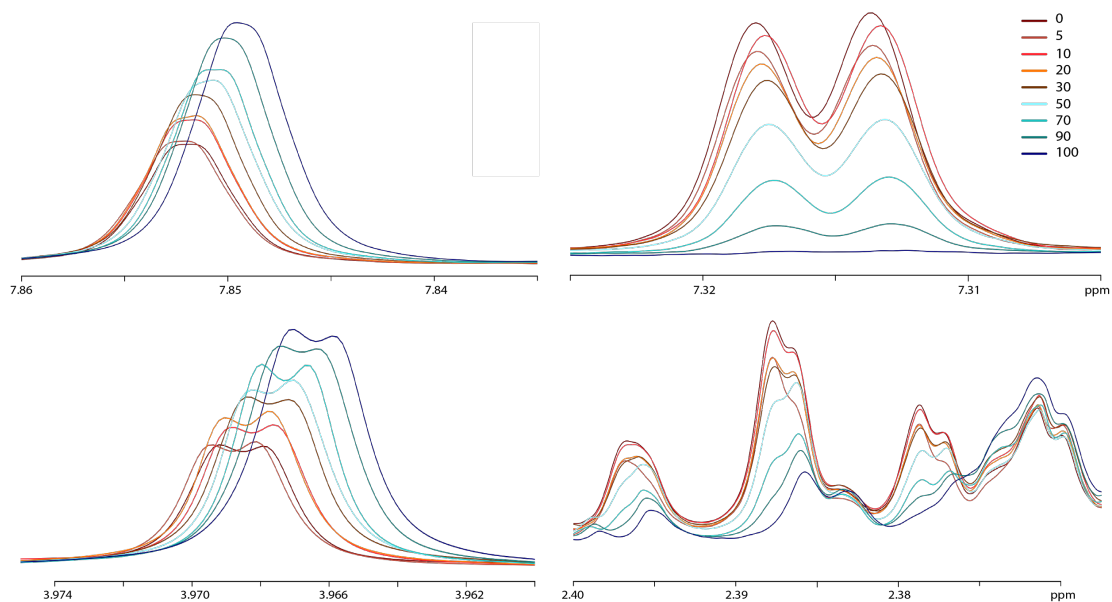


Figura 3.8. Perfil metabolómico ^1H -RMN del extracto metanólico de café Arábica mezclado con café Robusta en presentación tostado. En las señales 7.85 y 3.96 ppm correspondientes a la cafeína presenta un desplazamiento hacia campo alto y aumento de la intensidad de la señal a medida que aumenta el contenido de café Robusta desde 0 (rojo oscuro) hasta 100 % (azul oscuro), mientras que en las señales 7.32 y 2.37-2.40 ppm presenta disminución en la intensidad.

Los perfiles metabolómicos obtenidos para las dos especies muestran las grandes bondades que tiene la RMN, ya que sin necesidad de realizar una caracterización de todas las señales del espectro se podría mediante una inspección visual predecir a que especie pertenece o si existe algún tipo de adulteración, ya que los cambios son fácilmente observables. Sin embargo, hay que tener en cuenta que esto es posible siempre y cuando los espectros obtenidos sean de una excelente calidad. Por otro lado, se vuelve más complejo cuando se tienen los perfiles de café arábigo Colombiano y café arábigo de otros orígenes (Figura 3.9), ya que las diferencias entre ellos no son fácilmente reconocidas con una inspección visual y es necesario realizar un análisis más exhaustivo.

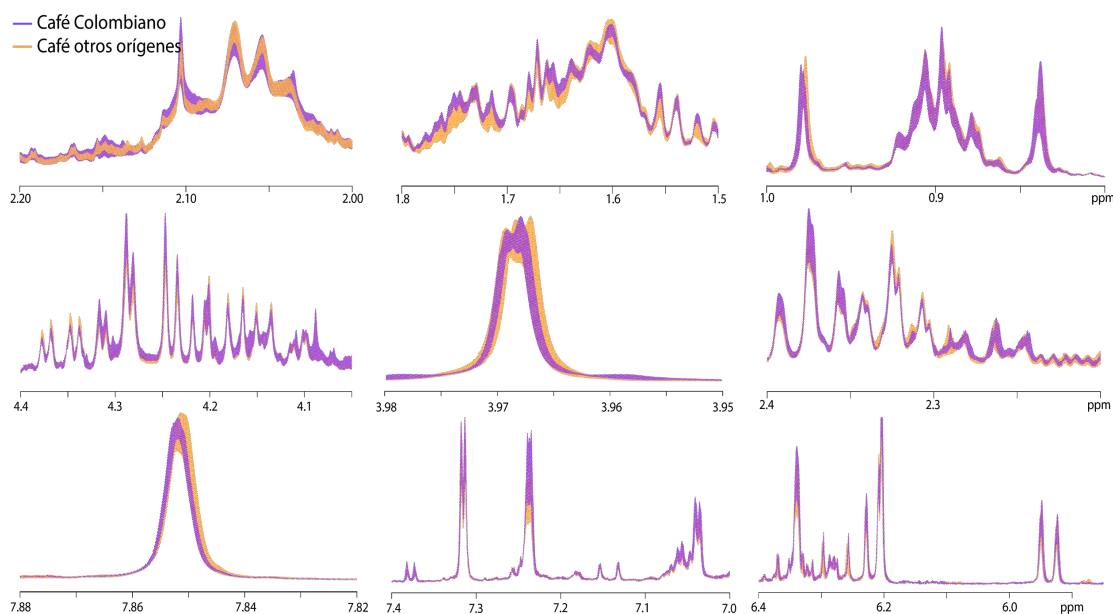


Figura 3.9. Perfil metabolómico ^1H -RMN de café arábigo Colombiano (violeta) y café arábigo de otros orígenes (naranja) en extractos metanólicos de semillas tostadas.

Así mismo se pueden observar cambios en los perfiles metabolómicos según el procesamiento del café. En la Figura 3.10 se presenta el caso de café descafeinado, el contenido de cafeína por lo general se reduce de 1 – 2 % a 0.02 - 0.3 % (Toci, Farah & Trugo, 2006; Farah, de Paulis, Trugo & Martin, 2006a). La pérdida de otros componentes del café depende del método de descafeinado. Mientras que el agua

tiende a eliminar muchos componentes solubles, la afinidad de los diversos disolventes orgánicos utilizados tiende a ser más específico. Los análisis de semillas descafeinados han demostrado que su composición química cambia considerablemente durante el proceso.

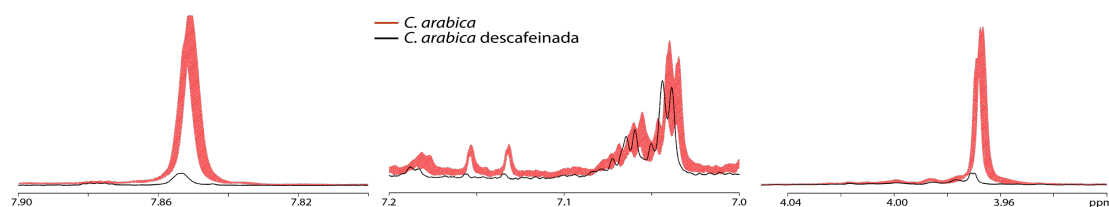


Figura 3.10. Perfil metabolómico ¹H-RMN de la especie *C. arabica* (rojo) y *C. arabica* descafeinada (negro) en extractos metanólicos de semillas de café tostado. Se evidencia la ausencia de la señal de cafeína en las regiones 3.97 y 7.85 ppm, y cambios en la región 7.0-7.2 ppm.

Aunque en el presente estudio no se realizó un modelo teniendo en cuenta muestras de café descafeinadas, los perfiles de estas muestras fueron de gran importancia cuando ALMACAFÉ entregaba una muestra que se consideraba sospechosa de adulteración, ya que se logró identificar algunas de estas muestras como descafeinadas con una inspección visual del espectro, sin embargo el modelo obtenido no permitió determinar el origen botánico y geográfico en estos casos.

3.3 ANÁLISIS NO SUPERVISADO

El PCA fue realizado para analizar los datos espectrales, con el fin de identificar y estudiar cualquier muestra que presentaba un comportamiento atípico en relación a su grupo para decidir su inclusión o exclusión en el estudio. Para cada grupo de muestras se analizó el gráfico de puntuaciones y aquellos datos que se ubicaron por fuera de la elipse, la cual representa un intervalo de confianza del 95 % utilizando el estadístico T² de Hotelling, fueron considerados datos atípicos.

El conjunto total de datos espectrales (543), incluyó las presentaciones de café verde (116) y tostado (422). En el grupo de datos pertenecientes a extractos metanólicos de

semillas de café verde con un total de 116 datos espectrales (*C. arabica* (92) y *C. canephora* (24)) se identificó a partir de un primer PCA (Figura 3.11A) la presencia de 3 datos (AC1337_BRAAS10080, AC1242_NUNAH12032 y AC1347_GUAAH12165) fuera del intervalo de confianza mostrando un comportamiento estadístico diferente al grupo de *C. arabica*, y 1 dato (AC1259_UGARH11076) fuera del intervalo de confianza para el grupo de *C. canephora*. Estos datos fueron excluidos de la matriz de datos espectrales y se realizó de nuevo un PCA (Figura 3.11B) en la cual se identificaron dos datos (AC1357_GUARH12176 y AC1292_BRARS11103) para el grupo *C. canephora* fuera del intervalo de confianza. Estos dos datos también fueron excluidos de la matriz de datos espectrales. Finalmente, se realizó un PCA con todos los datos incluidos en el grupo de semillas de café verde (110 datos espectrales, *C. arabica* (89) y *C. canephora* (21)) para evaluar si existía alguna tendencia debida al origen botánico (Figura 3.11C), fue posible observar una tendencia de agrupamiento de los datos en el primer componente para las dos especies en estudio, *C. arabica* (círculos rojos) y *C. canephora* (círculos azules).

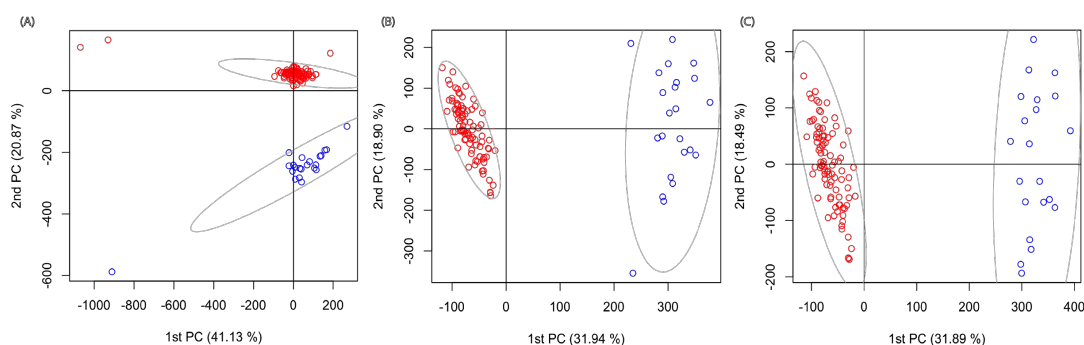


Figura 3.11. Gráfico de puntuaciones del PCA de (A) todos los datos espectrales de extractos metanólicos de semillas verdes de *C. arabica* (círculos rojos) y *C. canephora* (círculos azules), (B) los datos espectrales de extractos metanólicos de semillas verdes de *C. arabica* (círculos rojos) y *C. canephora* (círculos azules) excluyendo los datos atípicos obtenidos en (A), (C) Los datos espectrales de extractos metanólicos de semillas verdes de *C. arabica* (círculos rojos) y *C. canephora* (círculos azules) excluyendo los datos atípicos obtenidos en (B). Los datos situados por fuera de la elipse que señala el 95 % del intervalo de confianza son identificados como datos atípicos.

Para el grupo de datos pertenecientes a extractos metanólicos de semillas de café tostado con un total de 413 datos espectrales (*C. arabica* (266), *C. canephora* (46) y mezclas entre especies (101)) se identificó en el gráfico de puntuaciones del PCA para las especies puras (Figura 3.12A) 10 datos atípicos pertenecientes al grupo de *C. arabica* (AC1607_BRALA13337, AC1680_CTIAH14388, AC1229_PANAH12122, AC1625_CENAH13355 (2 repeticiones), AC1396_KENAH12208, AC1500_CRIANN13307, AC1213_BRAAS12064, AC1406_BRAA12215 (2 repeticiones)), y 2 datos atípicos pertenecientes al grupo de *C. canephora* (AC1423_MEXR_12233 y AC1350_CHIRS_12168 (2 repeticiones)). Estos datos fueron excluidos de la matriz de datos espectrales. Se destacó con este análisis que se identificaron 7 datos atípicos más con respecto a los identificados con el gráfico de puntuaciones del PCA de la Figura 3.11. Para el caso de las mezclas, en el gráfico de puntuaciones del PCA presentado en la Figura 3.12B se observaron 6 datos atípicos que corresponden a las muestras UV1220_M5_90, UV1206_M3_90, UV1234_M6_70, UV1227_M6_90, UV1225_M6_50, UV1235_M8_5. Finalmente, se realizó un PCA con todos los datos incluidos en el grupo de semillas de café tostado (395 datos espectrales, *C. arabica* (256), *C. canephora* (44) y mezclas entre especies (95)) para evaluar la tendencia debida al origen botánico, como se aprecia en la Figura 3.12C, fue posible observar una tendencia de agrupamiento de los datos en el segundo componente para las dos especies puras, ubicándose la especie *C. arabica* hacia el lado negativo y *C. canephora* hacia el lado positivo, además se observó que a medida que aumenta el porcentaje de *C. canephora* en la mezcla, las muestras se van desplazando del lado positivo hacia el negativo.

De igual forma se continuó con la exploración de los datos espectrales, en este caso con los correspondientes al grupo de *C. arabica* para observar si existía agrupamiento entre los cafés Colombianos y los cafés de otros orígenes. El PCA (Figura 3.13A) para los datos de los extractos metanólicos de semillas de café verde con un total de 89 datos espectrales (café Colombiano (48) y café de otros orígenes (41)) no presentaron agrupamiento. Mientras que el PCA (Figura 3.13B) para los datos de los

extractos metanólicos de semillas de café tostado con un total de 256 datos espectrales (café Colombiano (120) y café de otros orígenes (136)) se identificaron 5 datos atípicos correspondientes al grupo de café Colombiano (AC1212_MSMAH12043, AC1356_CPIAH12175, AC1465_CTIAH13296, AC1448_MCIAH13257, AC1640_CTIAH13373) que al ser eliminados del conjunto de datos se observó un agrupamiento de los dos grupos aunque superpuestos (Figura 3.13C).

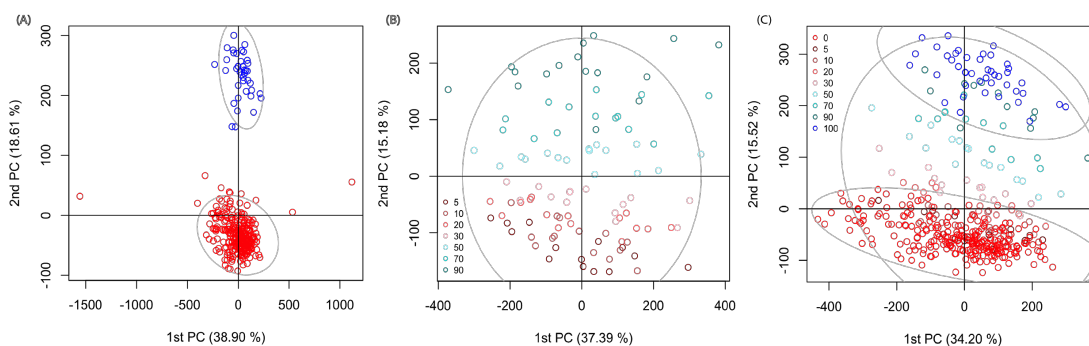


Figura 3.12. Gráfico de puntuaciones del PCA de (A) todos los datos espectrales de extractos metanólicos de semillas tostadas de *C. arabica* (círculos rojos) y *C. canephora* (círculos azules), (B) los datos espectrales de extractos metanólicos de las semillas tostadas de *C. arabica* mezcladas desde 5 hasta 90 % con *C. canephora* (círculos de rojo a azul cuando aumenta concentración de *C. canephora*), (C) los datos espectrales de extractos metanólicos de semillas tostadas de *C. arabica* (círculos rojos), *C. canephora* (círculos azules) y mezclas excluyendo los datos atípicos obtenidos en (A) y (B). Los datos situados por fuera de la elipse que señala el 95 % del intervalo de confianza son identificados como datos atípicos.

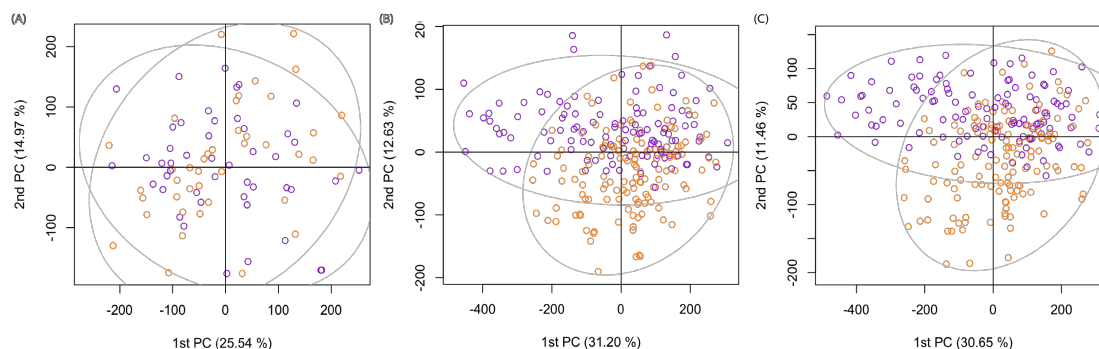


Figura 3.13. Gráfico de puntuaciones del PCA de (A) los datos espectrales de extractos metanólicos de semillas verdes de *C. arabica* de origen Colombiano (círculos violetas) y otros orígenes (círculos naranjas), (B) los datos espectrales de extractos metanólicos de semillas tostadas de *C. arabica* de

origen Colombiano (círculos violetas) y otros orígenes (círculos naranjas) y (C) los datos espectrales de (B) excluyendo los datos atípicos del análisis. Los datos situados por fuera de la elipse que señala el 95 % del intervalo de confianza son identificados como datos atípicos.

Con el objeto de comprender el comportamiento atípico de los datos espectrales que se excluyeron del estudio, la revisión de los espectros ^1H -RMN presentó una clara evidencia de su pobre calidad, por lo general las causas correspondieron a mala supresión de la señal del solvente, fallas en el shim y en la corrección de la línea base. En la Figura 3.14 se presenta un ejemplo donde se comparó un espectro de una muestra representativa de café verde arábigo (negro) con el espectro de uno de los datos atípicos correspondiente a la muestra de café Arábica AC1337_BRAAS10080 (rojo), se observó que existen diferencias entre los dos espectros en todas las regiones del espectro, así como una mala supresión del solvente (3.2 - 3.5 ppm).

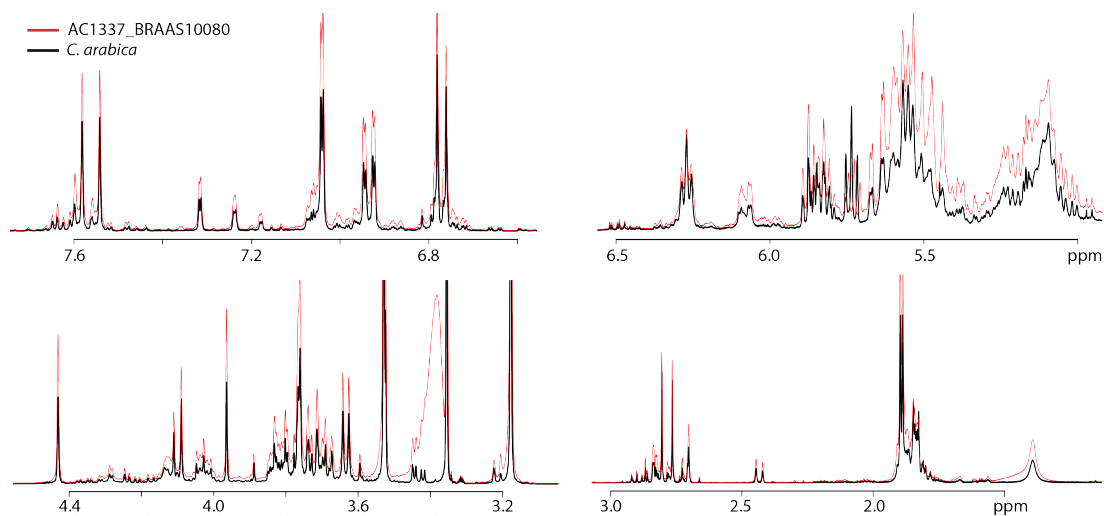


Figura 3.14. Comparación de los espectros ^1H -RMN de la muestra de café Arábica AC1337_BRAAS10080 (rojo) que resultó del PCA como dato atípico con una muestra representativa del grupo *C. arabica* (negro).

3.4 ANÁLISIS SUPERVISADO

3.4.1 MODELO DE CLASIFICACIÓN

Una vez definidas las muestras que presentaron buena calidad en los espectros, se procedió a realizar el análisis supervisado (PLS-DA). Como se señaló en el capítulo

anterior, este análisis utilizó el conocimiento previo que se tenía de las muestras para construir modelos estadísticos que permitieran, posteriormente, identificar cuáles eran las variables que mejor predecían la clase a la que pertenecía cada muestra.

El análisis discriminante supervisado se aplicó a una matriz de 500 datos espectrales (incluyendo repeticiones) x 1610 variables, de los cuales 110 datos correspondían a café verde y 390 a café tostado. La descripción de las muestras usadas para la construcción del modelo estadístico de clasificación se encuentra en el Anexo 6.1. A partir de estos datos espectrales, 436 datos espectrales (96, café verde y 340, café tostado) se utilizaron para construir y optimizar el modelo estadístico de clasificación usando PLS-DA (conjunto de entrenamiento y conjunto de prueba), mientras que los 64 datos restantes (14, café verde y 50, café tostado) conformaron el conjunto de validación externa. 100 modelos diferentes ($N = 100$) fueron construidos como una estrategia de doble validación cruzada y se calcularon los valores medios de los parámetros de validación S , Sp , R^2 , Q^2 , $MSEC$, $MSECV$ y $MSEEP$. En la Figura 3.15 se presenta el bucle de doble validación cruzada usado en el modelo estadístico de clasificación de origen botánico y geográfico para café verde y tostado.

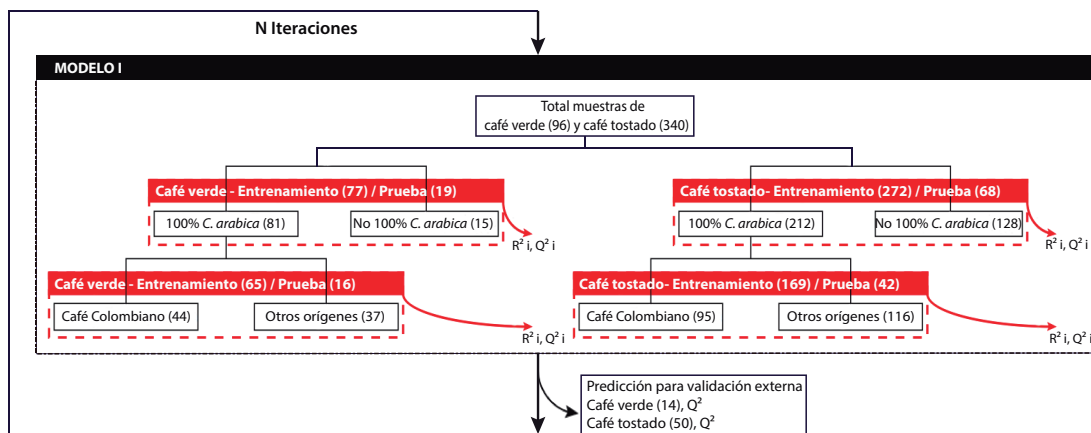


Figura 3.15. Representación del bucle de doble validación cruzada usado en el modelo estadístico de clasificación de origen botánico y geográfico para café verde y tostado.

Para café verde, el modelo resultante destacó una excelente separación de las muestras de acuerdo a su origen botánico, “*C. arabica*” vs “No *C. arabica*”,

empleando 12 variables latentes se obtuvieron valores medios para los conjuntos de entrenamiento y de prueba de S y Sp del 100 %, de R^2 y Q^2 de 1.0. Igualmente, el conjunto de validación externa obtuvo un $Q^2= 1.0$, en la Tabla A6.9a del Anexo 6.9 se presentan las predicciones y frecuencias (cantidad de veces que los 100 modelos estadísticos clasificaban correctamente las muestras), mostrando que los 100 modelos estadísticos predecían los mismos resultados. En este caso la clasificación fue perfecta debido a que el conjunto de datos no contenía mezclas de las especies, tratándose sólo de las especies puras. Para el caso de la clasificación por origen geográfico, “café Colombiano” vs “otros orígenes”, igualmente se obtuvo una excelente clasificación. Empleando 18 variables latentes se lograron valores medios para los conjuntos de entrenamiento y de prueba de S y Sp por encima del 97 %, de 1.0 y 0.873 para R^2 y Q^2 , respectivamente. Para el conjunto de validación externa se obtuvo un valor de $Q^2= 1.0$. Los resultados de las predicciones y frecuencias para estas muestras se encuentran en la Tabla A6.9a del Anexo 6.9. Los valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café verde se presentan en la Tabla 3.1.

Tabla 3.1. Valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café verde.

	“C. arabica” vs “No C. arabica”			“Café colombiano” vs “otros orígenes”		
	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validacion externa	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validacion externa
Número de muestras	77	19	6	65	16	8
Sensibilidad (%)	100	100	100	100	97.8	100
Especificidad(%)	100	100	100	100	96.3	100
R^2	1			1		
Q^2		1	1		0.873	1
SMEE	0			0		
SMECV		0			0.244	
SMEP			1			0

Además, se realizó la prueba de permutación construyendo 100 modelos estadísticos PLS-DA de clasificación con permutación de las etiquetas de clases, el p -value

obtenido tanto para la clasificación botánica como geográfica fue de 0.01, reflejando la validez de los resultados obtenidos.

Para café tostado, la primera clasificación, "*C. arabica*" vs "No *C. arabica*", se emplearon 12 componentes, mientras que 18 componentes fueron necesarios para la segunda clasificación, "*café Colombiano*" vs "*otros orígenes*". Como era de esperar, las tasas más altas se obtuvieron para la primera clasificación, se clasificaron para los conjuntos de entrenamiento y de prueba con valores medios de S y Sp mayores al 96 % , así como R^2 y Q^2 medio de 0.952 y 0.893, respectivamente. Al realizar una inspección más detallada de los resultados obtenidos se reveló que algunas muestras con bajo contenido porcentual de café Robusta (< 20 %) fueron clasificadas incorrectamente por algunos de los modelos como café Arábica puro, reflejando una disminución en el valor de Q^2 . En el conjunto de validación externa se obtuvo un valor de $Q^2 = 1.0$, los resultados de las predicciones y frecuencias para estas muestras se encuentran en la Tabla A6.9b del Anexo 6.9, donde se observó que la mayoría de las predicciones fueron iguales para los 100 modelos con una frecuencia de 1.0, excepto para 2 muestras con contenido de café Robusta menor al 10 % que fueron clasificadas correctamente pero con una frecuencia de 0.78 y 0.85, lo que indicó que alrededor de 80 modelos las clasificaron como "No *C. arabica*". Por lo tanto, es importante contar con un gran número de modelos (N=100) para evaluar la calidad de las predicciones, ya que reportar el resultado de un solo modelo podría generar conclusiones incorrectas.

Durante la segunda clasificación, "*café Colombiano*" vs "*otros orígenes*", las muestras se clasificaron con S y Sp media mayores a 93 % y con R^2 y Q^2 medio de 0.933 y 0.743. Estos resultados confirman anteriores trabajos sobre discriminación de origen de café usando NIRS (Haiduc, Gancel, & Leloup, 2006) y GC-MS (Risticvic, Carasek, & Pawliszyn, 2008). La mayoría de los errores de clasificación se produjeron entre cafés del mismo continente. Los cafés de Brasil se clasificaron generalmente correctamente, sobre todo aquellos sometidos a un proceso de post

cosecha diferente, proceso “seco” (también llamado método natural). Por el contrario, las muestras de Brasil preparadas a través de proceso “húmedo” fueron clasificadas erróneamente con más frecuencia, lo que sugiere un proceso post cosecha similar al llevado a cabo en Colombia. La dependencia de la composición química del grano de café con los procesos posteriores a la cosecha ya ha sido demostrada por Duarte y colaboradores (2010). En el conjunto de validación externa se obtuvo un valor de $Q^2 = 0.876$, los resultados de las predicciones y frecuencias para estas muestras se encuentran en la Tabla A6.9b del Anexo 6.9, donde se observó que dos muestras de Colombia fueron clasificadas incorrectamente con una frecuencia de clasificación de 0.81 y 0.84, esto quiere decir que la mayoría de los modelos (~80) las clasificaron como “*otros orígenes*”, mientras que una muestra de otro origen fue clasificada incorrectamente con una frecuencia de 0.67. Por lo tanto, es importante siempre reportar la predicción junto con la frecuencia de clasificación, en lugar de reportar una sola predicción usando el “mejor” modelo. Los valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café tostado se presentan en la Tabla 3.2.

Tabla 3.2. Valores medios de los parámetros de validación para los 100 modelos de clasificación por origen botánico y geográfico de las muestras de café tostado.

	“C. arabica” vs “No C. arabica”			“Café colombiano” vs “otros orígenes”		
	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validacion externa	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validacion externa
Número de muestras	272	68	50	169	42	40
Sensibilidad (%)	99.9	99.3	100	98.6	93.4	95.3
Especificidad(%)	97.4	96.9	100	98.1	94.0	99.0
R^2	0.952			0.933		
Q^2		0.893	1		0.743	0.876
SMEE	0.211			0.248		
SMECV		0.278			0.452	
SMEP			0			0.632

Se realizó la prueba de permutación para las muestras de café tostado igual como se realizó en café verde y se obtuvo el *p-value* para la clasificación botánica y

geográfica por debajo del valor mínimo (<0.05), confirmando que la clasificaciones son significativamente mejores y validando los resultados obtenidos.

ALMACAFÉ entregó 35 muestras de café tostado catalogadas como sospechosas de adulteración para ser evaluadas con los modelos de clasificación PLS-DA validados. Inicialmente, en una inspección visual de los espectros ^1H -RMN obtenidos de los extractos metanólicos se evidenció que 9 muestras presentaron ausencia de las señales de cafeína, ver Figura A6.9. Estas muestras fueron consideradas muestras descafeinadas y por tal razón no se realizaron las predicciones. Las 26 muestras restantes fueron evaluadas con los modelos estadísticos de clasificación de origen botánico y geográfico PLS-DA, los resultados de las predicciones y frecuencias se presentan en la Tabla A6.9c del Anexo 6.9. Se obtuvieron 8 muestras que se clasificaron como “No *C. arabica*” y 18 como “*C. arabica*”. Las 18 muestras clasificadas como “*C. arabica*”, 7 muestras se clasificaron como “*Café Colombiano*” y 11 como “*otros orígenes*”.

3.4.2 MODELO DE CUANTIFICACIÓN

Después de obtener los modelos de clasificación PLS-DA para origen botánico de las muestras de café tostado, se realizó un análisis de regresión PLS para cuantificar el contenido porcentual de café Robusta para aquellas muestras que se clasificaron como “No *C. arabica*”. Para este análisis se usaron 128 datos espectrales (30 *C. Arábica*, 30 *C. Robusta* y 95 mezclas entre las especies). El 80 % de los datos espectrales (103) fueron tomados para el conjunto de entrenamiento y el 20 % restante (25) para el conjunto de prueba. Se determinaron 12 variables latentes mediante la técnica de validación cruzada. Una gráfica del contenido porcentual de *C. Robusta* predicho vs observado se llevó a cabo (Figura 3.16), para el conjunto de entrenamiento se obtuvo un valor de R^2 de 0.9826 indicando una muy buena correlación lineal, este alto coeficiente de correlación lineal resultó comparable respecto a los valores obtenidos por otros autores que han cuantificado el contenido

de la variedad robusta por NIRS (Pizarro, Esteban-Díez, & González-Sáiz, 2007), ESI FT-ICR MS (Garret, Vaz, Hovell, Eberlin, & Rezende, 2012) y RMN (Cagliani *et al.*, 2013), y para el conjunto de prueba se obtuvo un valor de Q^2 de 0.9759.

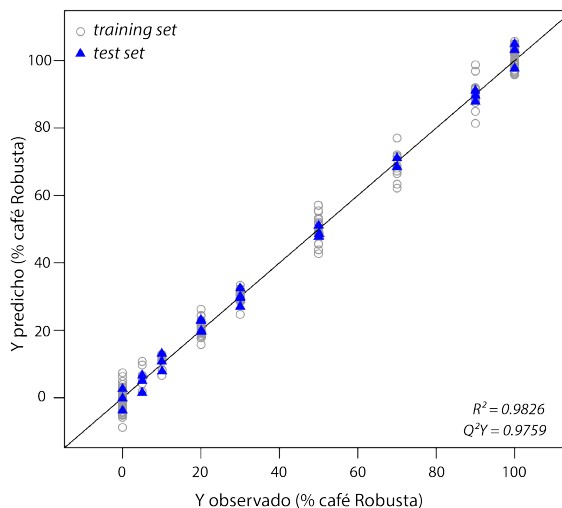


Figura 3.16. Y observado vs Y predicho del contenido porcentual de *C. Robusta* en muestras de *C. Arábica* tostadas obtenido para los conjuntos de entrenamiento y de prueba usando el modelo PLS.

En la Tabla A6.9c del Anexo 6.9 se presentan los resultados de las predicciones del contenido porcentual de café Robusta para las 8 muestras catalogadas como sospechosas de adulteración que fueron clasificadas por los modelos estadísticos de clasificación PLS-DA como “No *C. arabica*”. Se obtuvieron 2 muestras 100 % café Robusta y 6 muestras con contenido de mezcla con café Robusta entre el 40 y 95 %.

3.4.3 VARIABLES RELEVANTES

Una vez evaluada la significancia estadística de los modelos PLS-DA obtenidos, se realizó el análisis de los gráfico de puntuaciones y cargas correspondientes a cada uno de los modelos estadísticos para seleccionar las regiones del espectro (variables) que contribuyen de manera más significativa a la clasificación. El gráfico de cargas mostró cuánto contribuye cada una de las variables a los diferentes componentes principales, las variables ubicadas en los extremos son las que más contribuyeron a la separación de los grupos en el gráfico de puntuaciones.

En las Figuras 3.17 (izquierda) y 3.18 (izquierda) se presentan los gráficos de puntuaciones de uno de los modelos PLS-DA para la clasificación del origen botánico de muestras de café verde y tostado, respectivamente. En café verde se evidenció un mayor agrupamiento para cada grupo a diferencia de la mayor dispersión que se observó con el café tostado. Para entender estos resultados se debe tener en cuenta que las muestras de café tostado estaban conformadas por muestras que contenían mezcla de las dos especies y existía un mayor número de muestras. La contribución relativa de los *bins* o *buckets* (pequeñas regiones procesadas del espectro) en la agrupación de las muestras de café verde se representa en el gráfico de cargas de la Figura 3.17 (derecha), en donde se observó una gran cantidad de variables como {*bin* a 0.99 ppm} no identificada, {bins a 1.31, 1.32, 1.33, 1.34, 1.35, 1.36, 1.37 y 1.38 ppm} no identificadas – región ácidos grasos, {bins a 3.30, 3.45, 3.48, 3.49 y 3.55 ppm} no identificadas – región de la sacarosa, {bins a 3.61 y 3.62 ppm} no identificadas, {bins a 3.68, 3.69, 3.70 y 7.77 ppm} cafeína, {bins a 4.18 y 4.19 ppm} trigonelina, {*bin* a 3.83 ppm} no identificada, {bins a 5.1 y 5.2 ppm} no identificadas, {bins 6.15, 6.81 y 6.91 ppm} no identificadas – región ácidos clorogénicos.

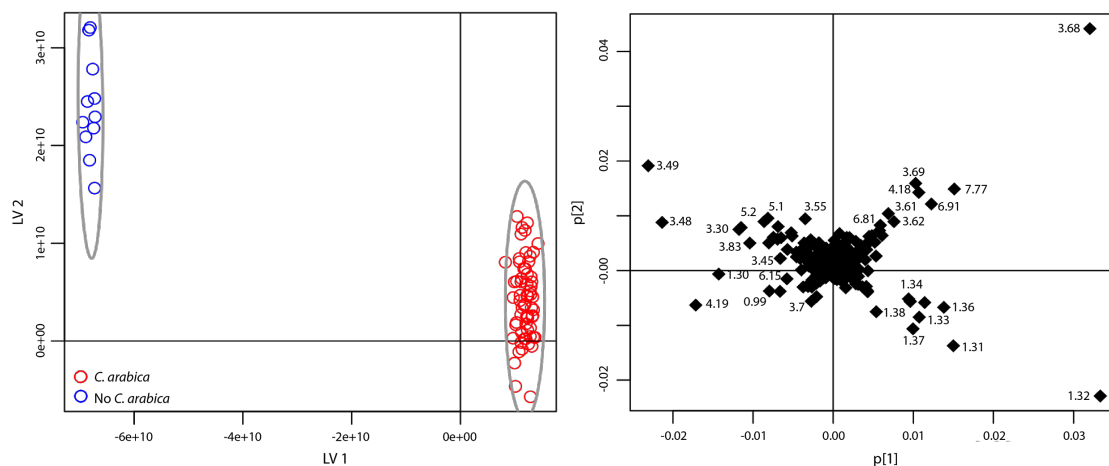


Figura 3.17. Gráfico de puntuaciones (izquierda) y cargas (derecha) correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen botánico entre “*C. arabica*” (círculos rojos) y “No *C. arabica*” (círculos azules) de café verde.

La contribución relativa de los *bins* o *buckets* en la agrupación de las muestras de café tostado se representa en el gráfico de cargas de la Figura 3.18 (derecha), en este caso se redujo el número de variables que contribuyeron a la clasificación, debido a que cambió el contenido de algunos compuestos químicos con el proceso de tostado, los polisacáridos disminuyeron, la sacarosa se degradó completamente, los ácidos orgánicos aumentaron y los ácidos clorogénicos disminuyeron (Flament, 2002; Murkovic & Derler, 2006). Las variables relevantes fueron {*bin* a 0.99 ppm} no identificada, {*bins* a 1.30, 1.31, 1.32 y 1.33 ppm} no identificadas – región ácidos grasos y {*bins* a 3.69, 3.70 y 7.77 ppm} cafeína, estas variables también fueron encontradas en café verde. Otras variables fueron {*bins* a 2.02, 2.03, 2.04 y 2.05 ppm} acetato y {*bins* a 2.17 y 2.20 ppm} no identificadas – región ácidos orgánicos y aminoácidos. La variable {*bin* a 0.99 ppm} fue muy importante para la diferenciación entre estas dos especies, ya que esta señal está presente en café Arábica más no en café Robusta, como se aprecia en la Figura 3.7.

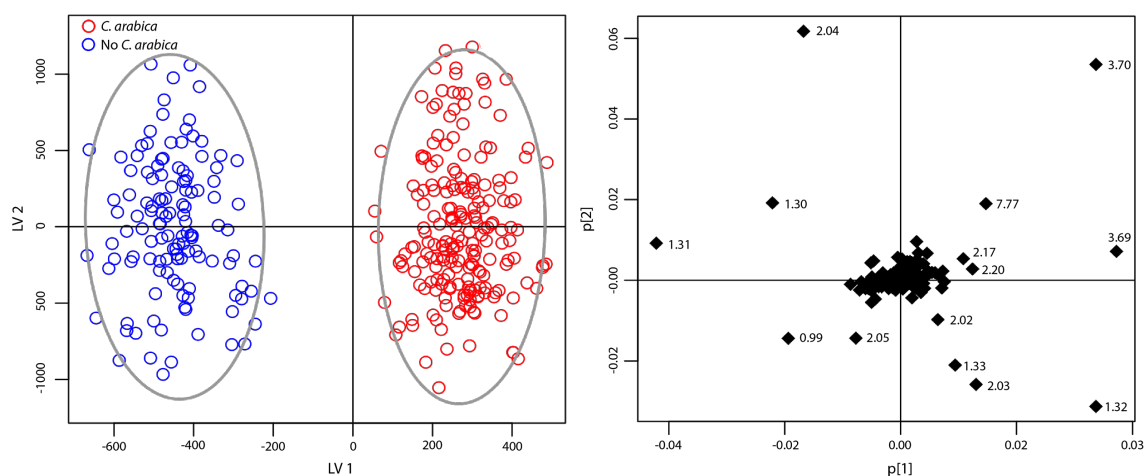


Figura 3.18. Gráfico de puntuaciones (izquierda) y cargas (derecha) correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen botánico entre “*C. arabica*” (círculos rojos) y “*No C. arabica*” (círculos azules) de café tostado.

En una comparación entre pares del origen botánico, “*C. arabica*” vs “*No C. arabica*”, usando el gráfico de cargas con escala de colores del modelo PLS-DA para muestras de café verde (Figura 3.19) se observó que la mayoría de las regiones del

espectro presentaron un alto potencial de discriminación entre las dos clases. Como era de esperar, en el gráfico de cargas con escala de colores del modelo PLS-DA para muestras de café tostado (Figura 3.20) se observaron menos variables que discriminaban, siendo las más importantes las regiones de la cafeína, *bin* 0.99 ppm y los ácidos clorogénicos, orgánicos y grasos.

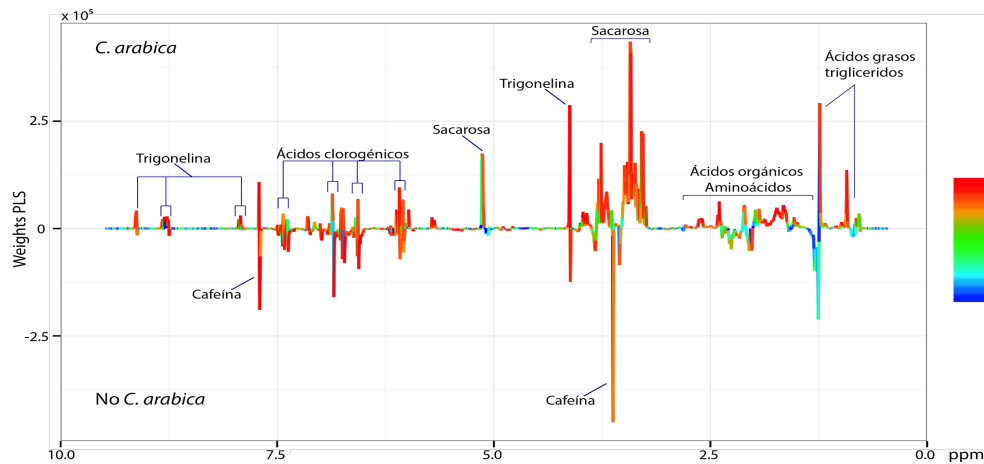


Figura 3.19. Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen botánico de café verde. El color rojo indica las variables de mayor peso discriminantes entre “*C. arabica*” y “No *C. arabica*”, mientras que el color azul indica las variables de menor peso discriminante.

En las Figuras 3.21 (izquierda) y 3.22 (izquierda) se presentan los gráficos de puntuaciones de uno de los modelos PLS-DA para la clasificación del origen geográfico de muestras de café verde y café tostado, respectivamente. Tanto en café verde como en café tostado se observó una clara separación entre los dos grupos, “*Café Colombiano*” vs “*otros orígenes*”. Se puede observar en la Figura 3.22 (izquierda) que las muestras de “*Café Colombiano*” se encontraron más agrupadas entre sí que las muestras de “*otros orígenes*”. La mayor variación entre muestras de “*otros orígenes*” se puede deber a la presencia de muestras de diferentes países y continentes con diferentes variedades de café, y diferentes condiciones climáticas, mientras que todos los cafés de Colombia tienen cierta similitud, en particular, la mayoría de los árboles de café son de la misma variedad.

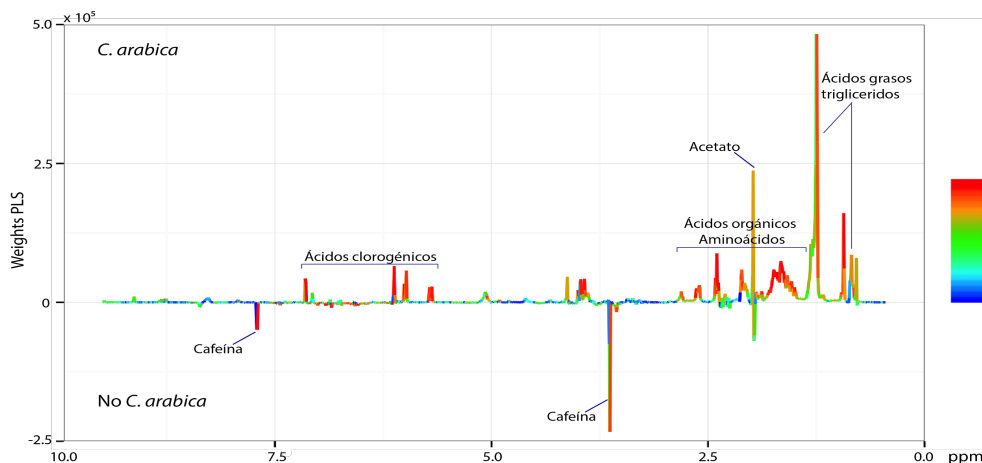


Figura 3.20. Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen botánico de café tostado. El color rojo indica las variables de mayor peso discriminantes entre “*C. arabica*” y “No *C. arabica*”, mientras que el color azul indica las variables de menor peso discriminante.

En el gráfico de cargas de la Figura 3.21 (derecha) se observó que los *bins* o *buckets* que contribuyeron a la agrupación de las muestras de café verde fueron {*bins* a 2.10, 2.16 y 2.30 ppm} no identificadas - región aminoácidos y ácidos orgánicos, {*bins* a 3.34, 3.35, 3.43, 3.44, 3.45, 3.46, 3.48, 3.49, 3.50, 3.53, 3.55 y 3.56 ppm} no identificadas - región de la sacarosa, {*bins* a 3.69 y 3.70 ppm} cafeína, {*bin* a 3.83 ppm} no identificada, {*bin* a 4.19 ppm} trigonelina, {*bins* a 5.18, 5.19 y 5.20 ppm} no identificadas, {*bins* a 6.11, 6.15, 6.63, 6.66, 6.80, 6.93, 7.45 y 7.49 ppm} no identificadas - región ácidos clorogénicos. En café tostado, las variables discriminantes se redujo debido a los cambios en los compuestos químicos debido al proceso de tostado, en el gráfico de cargas de la Figura 3.22 (derecha) se observó que los *bins* o *buckets* que contribuyeron a la clasificación fueron {*bin* a 0.99 ppm} no identificada, {*bins* a 1.30, 1.31, 1.32 y 1.36 ppm} no identificadas - región ácidos grasos, {*bins* a 2.02, 2.03, 2.04 y 2.05 ppm} acetato, {*bins* a 2.34 y 2.46 ppm} no identificadas - región ácidos orgánicos y aminoácidos, {*bins* a 3.69 y 3.70 ppm} cafeína, {*bin* a 4.19 ppm} no identificada, {*bins* a 6.19 y 7.21 ppm} no identificadas - región ácidos clorogénicos.

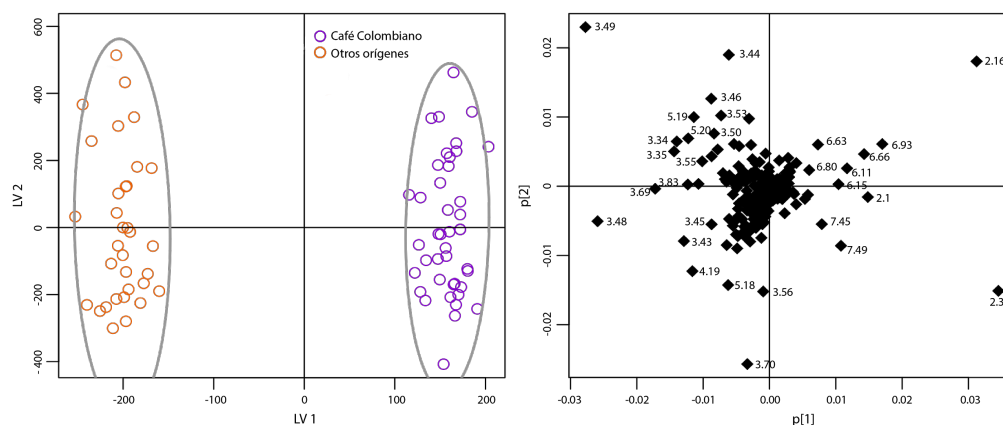


Figura 3.21. Gráficos de puntuaciones (izquierda) y cargas (derecha) correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen geográfico entre “Café Colombiano” (círculos violetas) y “otros orígenes” (círculos naranjas) de café verde.

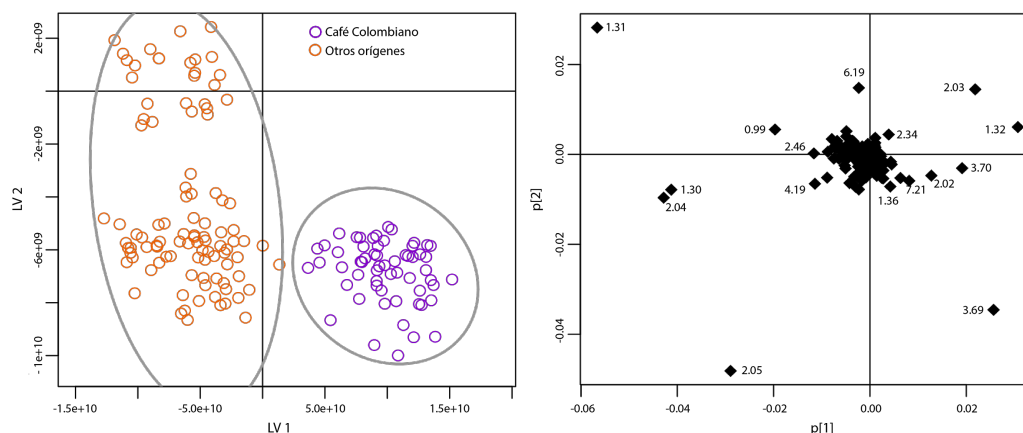


Figura 3.22. Gráfico de puntuaciones (izquierda) y cargas (derecha) correspondientes a uno de los 100 modelos PLS-DA para la clasificación del origen geográfico entre “Café Colombiano” (círculos violetas) y “otros orígenes” (círculos naranjas) de café tostado.

En el gráfico de cargas con escala de colores del modelo PLS-DA (Figura 3.23) para muestras de café verde se observó que la región de los ácidos clorogénicos, sacarosa y ácidos orgánicos discriminaban entre “café Colombiano” y “otros orígenes”. Mientras que en café tostado se observó en el gráfico de cargas con escala de colores del modelo PLS-DA (Figura 3.24) que las regiones del acetato y los ácidos clorogénicos, grasos y orgánicos, fueron discriminantes entre las dos clases.

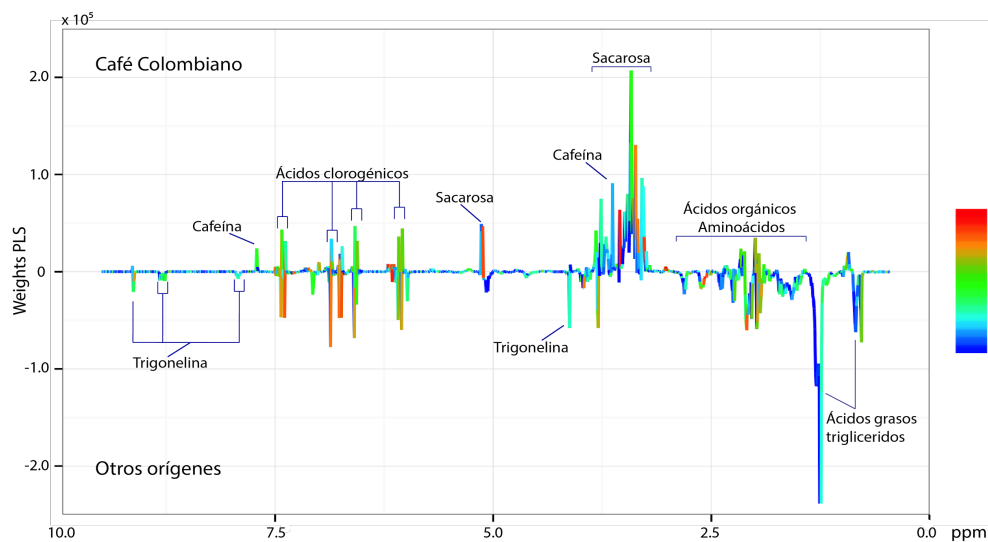


Figura 3.23. Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen geográfico de café verde. El color rojo indica las variables de mayor peso discriminantes entre “café Colombiano” y “otros orígenes”, mientras que el color azul indica las variables de menor peso discriminante.

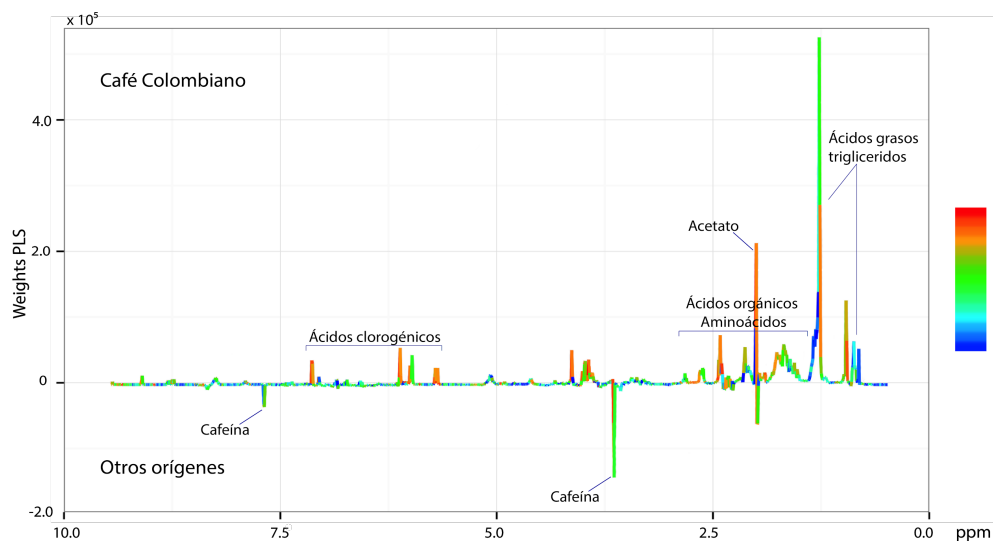


Figura 3.24. Gráfico de cargas con escala de colores obtenido del modelo PLS-DA para origen geográfico de café tostado. El color rojo indica las variables de mayor peso discriminantes entre “café Colombiano” y “otros orígenes”, mientras que el color azul indica las variables de menor peso discriminante.

3.5 EVALUACIÓN DEL EFECTO DEL GRADO DE TOSTADO SOBRE LOS MODELOS DE CLASIFICACIÓN

Una vez obtenidos los modelos de clasificación de origen botánico y geográfico, se procedió a evaluar el grado de tostado bajo dos temperaturas diferentes de procesamiento de los granos de café sobre la predicción del conjunto de validación externa, con el fin de determinar hasta que punto se podían usar los modelos sin incluir muestras con diferentes grados de tostado en el conjunto de entrenamiento.

En la Figura 3.25 se presenta la superposición de los espectros ^1H -RMN de los extractos metanólicos de una muestra de café a 4 diferentes grados de tostado (medio (22.5 L*), oscuro medio (21.0 L*), oscuro (17.0 L*) y muy oscuro (15.5 L*)) procesadas a 200 °C, se observó cambios en varias regiones del espectro, principalmente para la muestra con un grado de tostado muy oscuro (15.5 L*). Un perfil similar se obtuvo para los 4 diferentes grados de tostado de las muestras de café procesadas a 180 °C. Para el grado de tostado muy oscuro se evidenció cambios en la región del espectro entre 6.1 a 7.7 ppm correspondiente a los ácidos clorogénicos, con una reducción de las señales y con un aumento en la región de los ácidos alifáticos, principalmente entre 2.2 a 2.3 ppm. Esto se podría explicar debido a que los ácidos clorogénicos son severamente degradados durante el proceso de tostado, desapareciendo casi todo los 3 isómeros de los ácidos clorogénicos después del grado de tostado oscuro de acuerdo a lo reportado por Wei y colaboradores (Wei, Furihata, Koda, Hu, Miyakawa, & Tanokura, 2012b). Igualmente se han reportado incrementos en los ácidos alifáticos, principalmente del ácido quínico, cuando se aumenta el tiempo de tostado, esto se debe a la degradación de la glucosa, sacarosa y fructosa, y también de los ácidos clorogénicos (Wang & Lim, 2012).

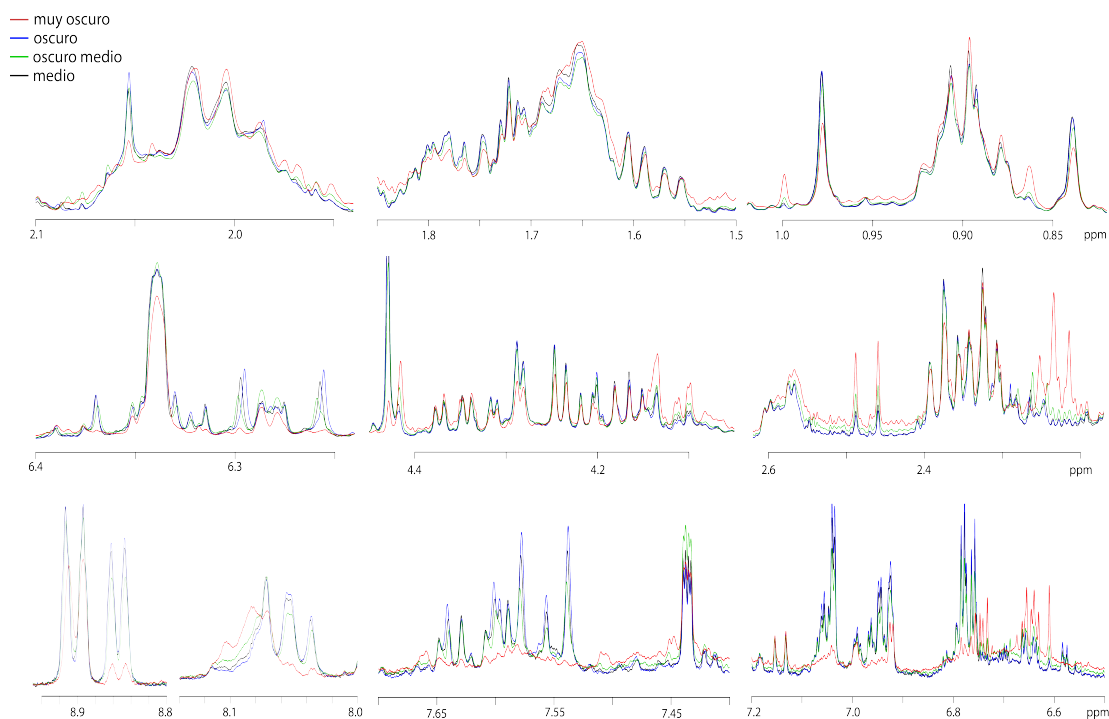


Figura 3.25. Superposición de los espectros ^1H -RMN de los extractos metanólicos de una muestra de café con 4 diferentes grados de tostado procesada a una temperatura de 200 °C. Grado de tostado: muy oscuro (rojo), oscuro (azul), oscuro medio (verde) y medio (negro).

Se realizó PCA con el fin de verificar la posibilidad de discriminación entre los diferentes grados de tostado y las temperaturas usadas en el proceso. En la Figura 3.26 se presentan los gráficos de puntuaciones obtenidos con el PCA de los datos espectrales de los 4 diferentes grados de tostado para las muestras de café procesadas a 200 °C (A) y a 180 °C (B). Se aprecia en las Figuras 3.26(A) para la temperatura de 200 °C, que las muestras con grado de tostado medio, oscuro medio y oscuro se agrupan en la parte positiva del PC1 y PC2, mientras que las muestras con grado de tostado muy oscuro se agrupan muy alejadas al resto de las muestras en la parte negativa del PC1 y PC2, esto quiere decir que el grado de tostado muy oscuro no se caracteriza por ninguna de las variables de los demás grados de tostado, tomando valores bajos para todas las variables en los dos componentes. En la Figura 3.26(B) se conserva el agrupamiento alejado de las muestras con grado de tostado muy oscuro con respecto a los otros grados de tostado, pero en este caso se agrupan en la parte

positiva del PC2, las variables del grado de tostado muy oscuro se correlacionan de forma negativa con los grados de tostado medio y oscuro medio, y de forma positiva con algunas variables del grado de tostado oscuro. Con este análisis se confirmó que las diferencias observadas en los espectros de las muestras con un grado de tostado muy oscuro se ven reflejadas en el agrupamiento diferente con respecto a los demás grados de tostado. Por otra parte, se realizó el PCA para todo el conjunto de datos espectrales obtenidos con temperatura de procesamiento de 200 °C y 180 °C, y como se observa en la Figura 3.26C no hay ningún agrupamiento por temperatura, pero igualmente se observó la misma tendencia de agrupamiento de las muestras de grado de tostado muy oscuro.

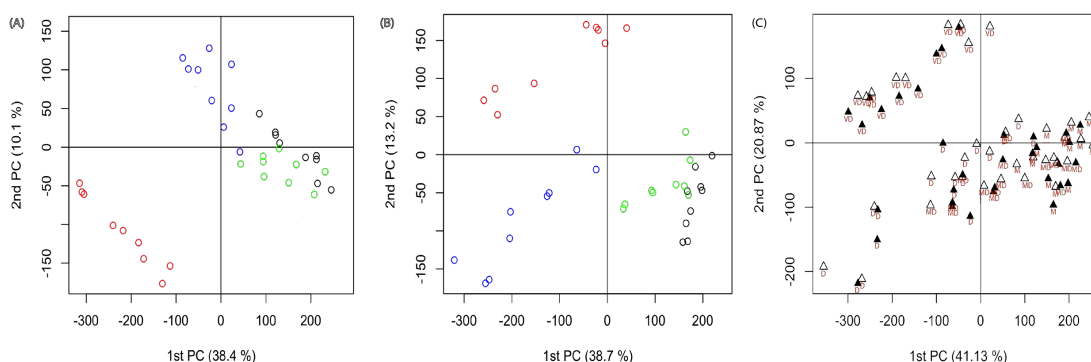


Figura 3.26. Gráfico de puntuaciones del PCA de los datos espectrales de extractos metanólicos de muestras de café tostado con diferentes grados de tostado a (A) 200 °C, (B) 180 °C, grado de tostado: muy oscuro (círculos rojos, VD), oscuro (círculos azules, D), oscuro medio (círculos verdes, MD) y medio (círculos negros, M) y (C) ambas temperaturas, 200 °C (triángulos negros) y 180 °C (triángulos blancos).

Se realizó PLS-DA sobre los datos de los extractos metanólicos de las semillas de café tostado para investigar los cambios generales en la composición del café al aumentar el grado de tostado con respecto al grado de tostado medio con el cual fue realizado el modelo de clasificación botánico y geográfico. El grado muy tostado fue el que presentó más diferencias con respecto al grado de tostado medio, se observó que la mayoría de los metabolitos son discriminantes asociados al proceso de tostado, la trigonelina y ácidos clorogénicos fueron mayores en grado de tostado medio, mientras que las regiones de los ácidos orgánicos y lípidos dieron los niveles más

altos para el grado muy tostado (Figura 3.27A). Los ácidos clorogénicos son severamente degradados durante el proceso de tostado, con pocos minutos de torrefacción hay una marcada disminución de 5-ACQ y un descenso estable en 4-ACQ y 3-ACQ (Wei *et al.*, 2012b). Casi todos los isómeros de ácido clorogénico desaparecieron después de tostado muy oscuro con tiempos mayores de 11 y 14 min para 200 y 180 °C, respectivamente, así como los niveles en la región de los ácidos orgánicos se incrementaron.

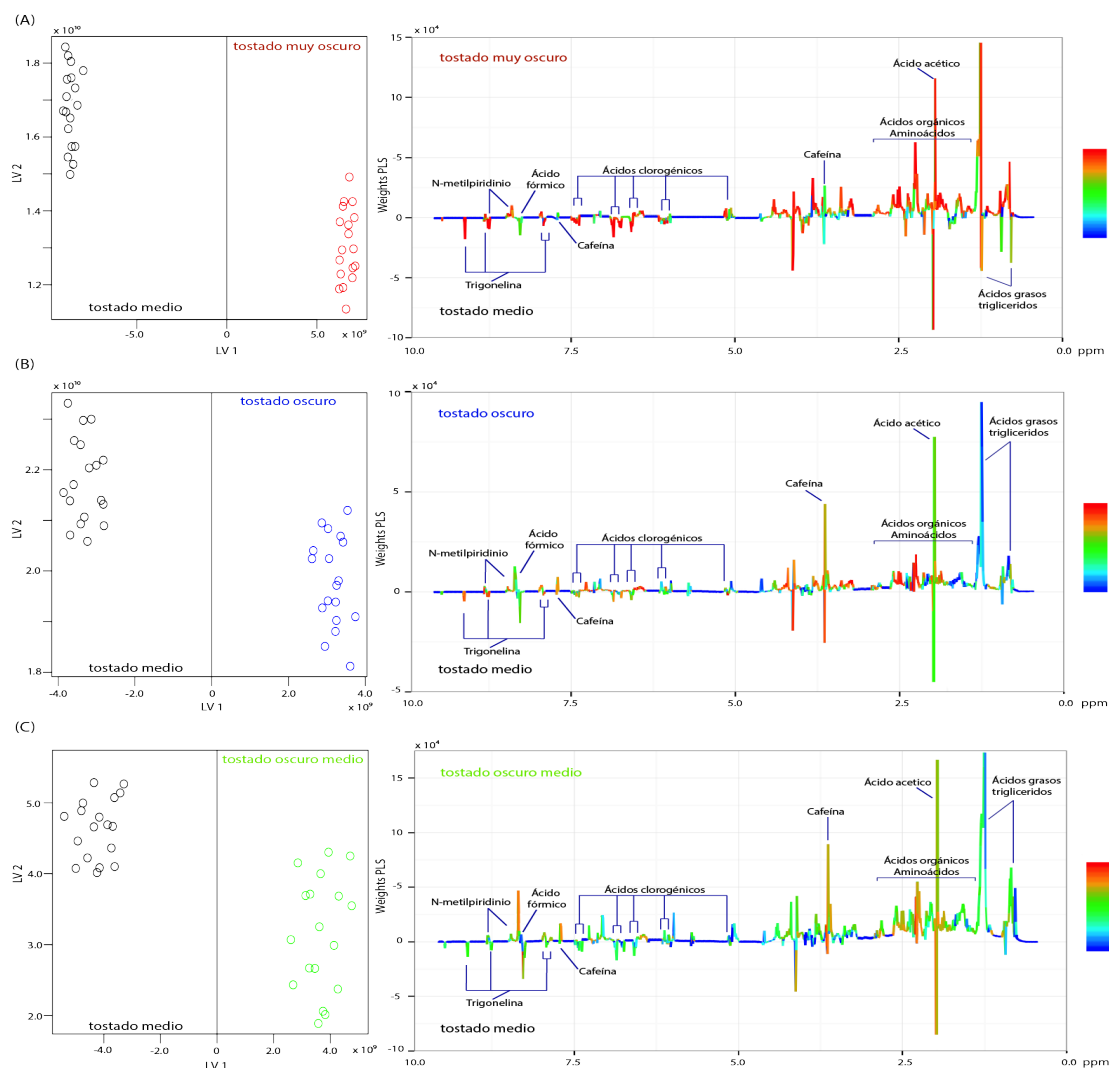


Figura 3.27. Gráfico de puntuaciones y cargas obtenidos del modelo PLS-DA para (A) grado de tostado medio vs muy tostado, (B) grado de tostado medio vs tostado oscuro (C) grado de tostado medio vs tostado oscuro medio.

Reportes confirman que N-metilpiridinio y ácido nicotínico son los principales productos no volátiles de la pirólisis de la trigonelina (Taguchi, Sakaguchi, & Shimabayashi, 1995; Stadler *et al.* 2002) y presentan una correlación positiva con el grado de tostado, de esta forma se observó que N-metilpiridinio iba aumentando a medida que aumentaba el grado de tostado (Figuras 3.27A-C) y tomó importancia como metabolito discriminante durante el proceso de tostado. Además, se observó en la Figura 3.27 que a medida que disminuyó el grado de tostado son menos el número de variables que contribuyeron a la diferenciación con el grado de tostado medio, solamente la región que continuó aportando a la diferenciación fue la de los ácidos orgánicos.

Teniendo en cuenta la anterior información se procedió a evaluar el conjunto de validación externa con los modelos de clasificación PLS-DA para origen botánico y geográfico. Como era de esperar las predicciones con los 4 grados de tostado para las dos temperaturas de tostado dieron valores de $Q^2 < 0.200$ tanto para origen botánico como geográfico. Se obtuvieron predicciones erróneas sobre el origen botánico: todas las muestras de café Arábica con grado de tostado muy oscuro fueron clasificadas como café Robusta. Por consiguiente, las muestras de café con grado de tostado muy oscuro fueron eliminadas del conjunto de validación externa y se evaluaron los grados de tostado oscuro, oscuro medio y medio para ambas temperaturas, en este caso se obtuvieron valores de Q^2 alrededor de 0.5 disminuyendo en comparación con 0.876 que se había obtenido con el grado de tostado medio (Tabla 3.3). Una forma de mejorar la predicción fue eliminando de la matriz de datos una pequeña región correspondiente a los ácidos orgánicos (*bins* 2.18 a 2.30 ppm), la cual se evidenció anteriormente que continuaba presentando diferencias entre los grados de tostado (Figuras 3.27B-C), logrando alcanzar valores de Q^2 de 0.756 y 0.803 para 200 °C y 180 °C, respectivamente (Tabla 3.3).

Tabla 3.3. Valores medios de los parámetros de validación para evaluar el efecto del proceso de tostado en la predicción del origen geográfico.

	Modelo de clasificación del origen geográfico			Efecto del proceso de tostado en la predicción del origen geográfico			
	Conjunto de entrenamiento	Conjunto de prueba	Conj. validacion externa	Conj. validacion externa	Conj. validacion externa*	Conj. validacion externa	Conj. validacion externa*
Número de muestras	169	42	40	40	40	40	40
Grado de tosti3n	M	M	M	M, MD, D	M, MD, D	M, MD, D	M, MD, D
Temperatura de tosti3n (°C)	200	200	200	200	200	180	180
Sensibilidad (%)	99.8	93.4	95.3	80.0	90.9	85.6	92.7
Especificidad (%)	99.6	94.0	99.0	85.0	100.0	87.0	100.0
R^2	0.985						
Q^2		0.743	0.876	0.523	0.756	0.584	0.803
SMEE	0.0758						
SMECV		0.452					
SMEP			0.632	0.170	0.894	1.370	0.742

* Eliminaci3n de la regi3n del espectro $^1\text{H-MNR}$ bins desde 2.18 a 2.30 ppm

M = tostado medio, MD = tostado oscuro medio, D = tostado oscuro

Reportes indican que los ácidos clorogénicos, trigonelina y ácidos orgánicos pueden ser indicadores del grado de tostado (Wei, *et al.*, 2012b), por tal motivo para aprovechar la gran informaci3n que contienen los datos espectrales de $^1\text{H-RMN}$ se realizó un análisis de regresi3n PLS para poder predecir el grado de tostado de las muestras antes de realizar el análisis de clasificaci3n con los modelos PLS-DA, con el fin de que las muestras con grados de tostado muy oscuro $< 17 \text{ L}^*$ no entrarán en las predicciones de los modelos. Para este análisis se usaron 78 datos espectrales (12 de 15.5 L^* , 16 de 17.0 L^* , 17 de 21.0 L^* , 16 de 22.5 L^* y 17 de 45.0 L^* , este último valor de color corresponde al café verde). El 80 % de los datos espectrales (62) fueron tomados para el conjunto de entrenamiento y el 20 % restante (16) para el conjunto de prueba. Se determinaron 8 variables latentes mediante la técnica de validaci3n cruzada. Una gráfica del grado de tostado predicho vs observado se llevó a cabo (Figura 3.28), para el conjunto de entrenamiento se obtuvo un valor de R^2 de 0.915 indicando una muy buena correlaci3n lineal y para el conjunto de prueba se obtuvo un valor de Q^2 de 0.898.

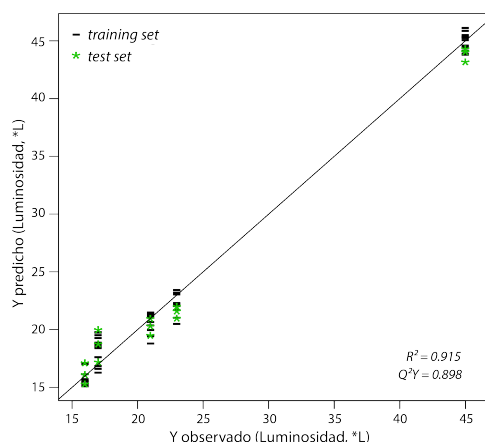


Figura 3.28. *Y predicho vs Y observado de la variable Luminosidad (°L) para predecir el grado de tostado del café obtenido para los conjuntos de entrenamiento y prueba usando el modelo PLS.*

3.6 EVALUACIÓN DE LAS TÉCNICAS ANALÍTICAS GC-C-IRMS, GC-MS Y ¹H-RMN PARA DETERMINAR EL ORIGEN GEOGRÁFICO DEL CAFÉ

La composición de isótopos estables, como $\delta^2\text{H}$, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$, determinado por IRMS (Serra *et al.*, 2005; Santato *et al.*, 2012), el análisis de compuestos volátiles y semi-volátiles por HS-SPME-GC-TOFMS (Risticvic, Carasek, & Pawliszyn, 2008), el análisis de aminoácidos y carbohidratos por LC-MS y GC-FID, respectivamente (Choi *et al.*, 2010), así como los estudios de ¹H-RMN (Consonni, Cagliani, & Cogliati, 2012) y ¹³C-RMN (Wei *et al.*, 2012) se han utilizado para la identificación del origen geográfico del café, presentando un alto potencial de estas técnicas para este propósito. Sin embargo, estos estudios raramente manejan muestras de café de los países vecinos. Aquí se evalúan las técnicas GC-C-IRMS, GC-MS y ¹H-RMN combinadas con quimiometría en muestras de café tostado (*Coffea arabica* L.) de tres países de América del Sur, con el objetivo de obtener una comparación de su capacidad para distinguir los cafés de Colombia, Brasil y Perú.

Los datos de la relación isotópica $\delta^{13}\text{C}$ de cafeína obtenidos por GC-C-IRMS y la composición porcentual (ítem 2.6) de siete compuestos: 2-furanmetanol, ácido

palmítico (C16:0), cafeína, dos tocoferoles (α y β) y dos esteroides (estigmasterol y sitosterol) obtenidos del perfil GC-MS se reportan en las Tablas 6.10a y 6.10b, respectivamente, del Anexo 6.10. Los espectros de huella metabólica basada en ^1H -RMN fueron procesados eliminando las señales correspondientes al metanol (δ 3.14-3.55 ppm) y TMS (δ -0.2 a 0.2 ppm), así como las regiones espectrales de los extremos. Luego se empleó el método de disminución de la resolución espectral (*binning*) con *bins* de anchura igual a δ 0.04 ppm para formar una matriz de 34 espectros (muestras) x 1610 variables.

3.6.1 ANÁLISIS POR GC-C-IRMS

En la Figura 3.29 se presenta la distribución de la relación de $\delta^{13}\text{C}$ de cafeína para los 3 países en estudio. El diagrama de caja permitió observar que no es posible discriminar entre Colombia y Brasil, ni entre Colombia y Perú utilizando esta variable, ya que las muestras se superponieron por dos o más cuartiles, es decir, por más de la mitad de las muestras.

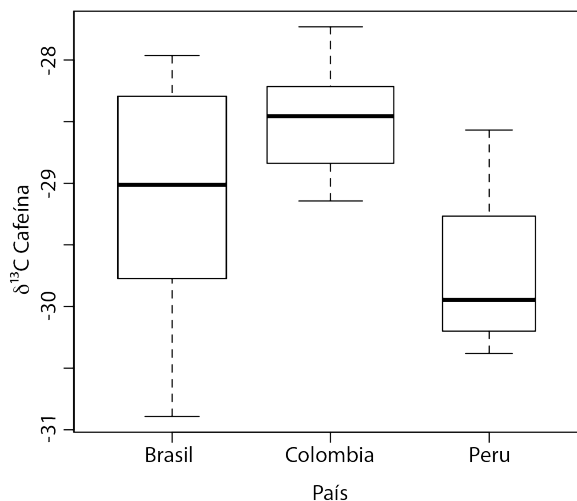


Figura 3.29. Diagrama de caja de los datos de la relación de $\delta^{13}\text{C}$ de cafeína obtenidos por GC-C-IRMS. Los valores de la relación isotópica para los países Colombia y Brasil presentan una gran superposición.

Este resultado coincidió con los resultados reportados por Weckerle y colaboradores

(2002) en donde presentan la superposición de los intervalos de valores de 3 isótopos estables diferentes para muestras de Brasil, Colombia y Costa Rica, utilizando un menor número de muestras. La relación de isótopos estables se ha asociado a las condiciones climáticas durante el crecimiento de plantas (Santato, Bertoldi, Perini, Camin, & Larcher, 2012; Rodrigues, Maia, Miranda, Ribeirinho, Nogueira, & Máguas, 2009) lo que ha permitido la discriminación cuando éstos son significativamente diferentes, esto podría no ser el caso de los países vecinos. De hecho, Serra y colaboradores (2005) destacan la similitud de las condiciones climáticas que se encuentran en las zonas seleccionadas para el cultivo del café en América del Sur.

3.6.2 ANÁLISIS POR GC-MS

Con el fin de evaluar el potencial de GC-MS para discriminar cafés de Colombia, Perú y Brasil, se realizó un análisis OPLS-DA utilizando estas tres clases. El 80 % de los datos espectrales (27) fueron tomados para el conjunto de entrenamiento para construir 100 modelos diferentes usando 2 LV y una estrategia de validación cruzada (Szymańska, Saccenti, Smilde, & Westerhuis, 2012; Smit, van Breemen, Hoefsloot, Smilde, Aerts, & de Koster, 2007) para predecir el origen del 20 % restante (7), conjunto de prueba. El promedio de los parámetros de calidad son: $S = 90.7 \%$, $Sp = 82.0 \%$, $R^2 = 0.80$ y $Q^2 = 0.57$. En la Figura 3.30(A) se presentan los gráficos de puntuaciones de uno de los mejores modelos obtenidos, confirmando visualmente la baja tendencia de separación y discriminación de las tres clases. En las Figuras 3.30(B-C) se presentan los mejores modelos obtenidos para la clasificación por pares, Colombia *vs* Perú y Colombia *vs* Brasil, respectivamente. Este análisis por separado señaló una diferencia significativa en el rendimiento de las dos clasificaciones, ya que se realizó una discriminación muy baja entre café Colombiano y Brasileño ($R^2 = 0.81$ y $Q^2 = 0.55$) con respecto a la discriminación entre café Colombiano y Peruano ($R^2 = 0.91$ y $Q^2 = 0.82$). A pesar de esto, los gráficos de cargas que se muestran en la Figura 3.30(C) permitieron identificar variables potencialmente relevantes, como: ácido palmítico (C16:0), β -tocoferol, sitosterol y estigmasterol, de acuerdo a trabajos

previos (Martín, Pablos, González, Valdenebro, León-Camacho, 2001; Romano, Santini, Grottaglie, Manzo, Visconti, Ritieni, 2014; Alves, Casal, Alves, Oliveira, 2009; González, Pablos, Martín, León-Camacho, Valdenebro, 2001) donde reportaron el papel de los ácidos grasos y tocoferoles para la discriminación del origen botánico de café, nuestros resultados mostraron que estos compuestos también podrían ser buenos candidatos para la discriminación del origen geográfico del café.

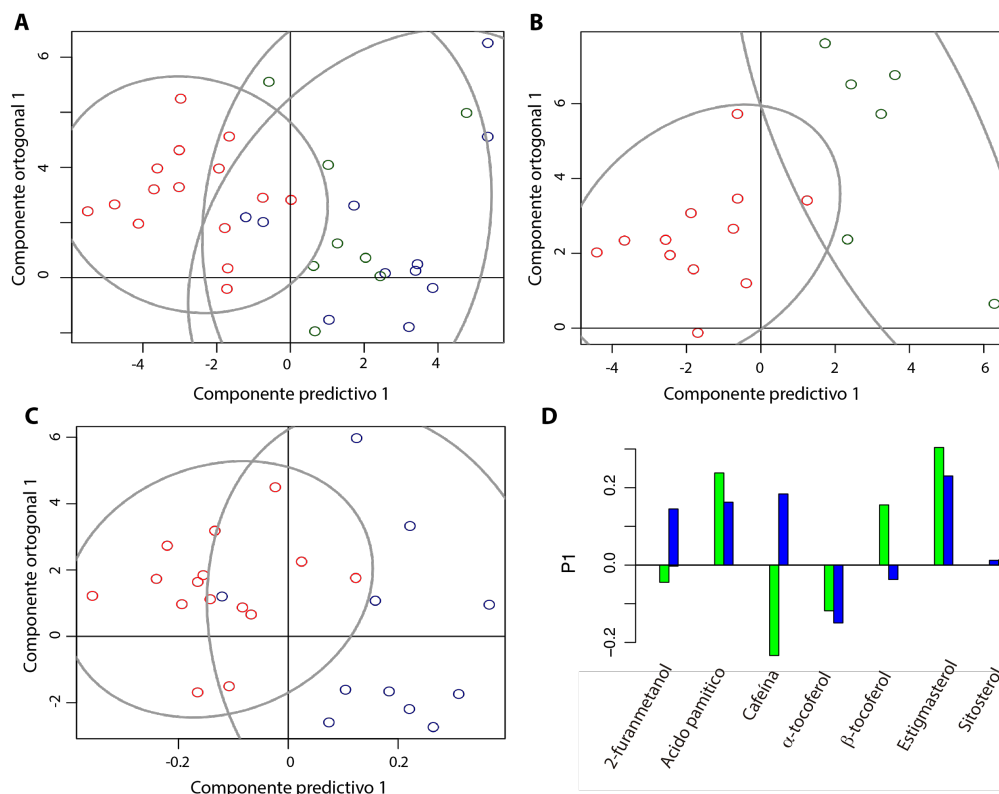


Figura 3.30. Gráfico de puntuaciones del análisis OPLS-DA de muestras de café obtenidos por GC-MS utilizando (A) cada país como una clase diferente, Colombia (círculos rojos), Perú (círculos verdes) y Brasil (círculos azules), (B) Colombia vs Perú, (C) Colombia vs Brasil, (D) Gráfico de cargas del análisis OPLS-DA obtenidas por GC-MS para los pares, Colombia vs Perú (barra verde) y Colombia vs Brasil (barra azul).

3.6.3 ANÁLISIS POR RMN

Un procedimiento similar como se acaba de reportar para los datos de GC-MS se aplicó a los datos de huella metabolómica ^1H -RMN. Se realizó un análisis OPLS-DA con todas las muestras utilizando los tres orígenes como clases. Una vez más, se

construyeron 100 modelos para discriminar los cafés de los tres orígenes, usando 8 variables latentes. El gráfico de puntuaciones de uno de los mejores modelos se muestra en la Figura 3.31. La simple inspección visual de los resultados reveló una mejor clasificación que el obtenido por GC-MS. Esto se confirmó por el promedio de los parámetros de calidad obtenidos: $R^2 = 0.85$ y $Q^2 = 0.78$.

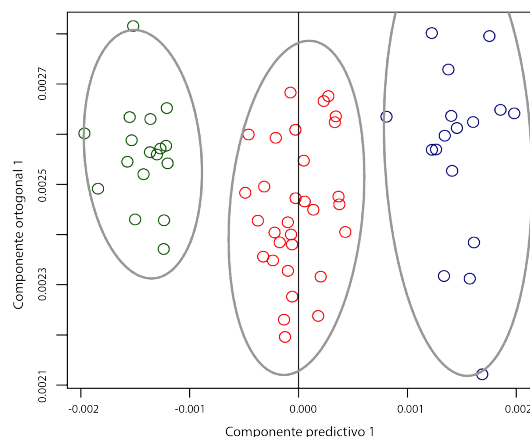


Figura 3.31. Gráfico de puntuaciones del análisis OPLS-DA para la clasificación de café con huella metabolómica ^1H -RMN de acuerdo a los tres países de origen: Colombia (círculos rojos), Brasil (círculos azules) y Perú (círculos verdes).

Se realizó el análisis por pares de países de manera similar que las indicadas para GC-MS, los gráficos de puntuaciones para uno de los 100 modelos OPLS-DA se muestran en la Figura 3.32(A y C) para Colombia vs Brasil y Colombia vs Perú, respectivamente. La huella metabolómica ^1H -RMN permitió una discriminación más precisa en comparación con GC-MS, que se tradujo en mayores valores de R^2 y Q^2 . De hecho, tanto la separación entre los cafés de Colombianos y Brasileños ($R^2 = 1.00$, $Q^2 = 0.98$) y entre los cafés Colombianos y Peruanos ($R^2 = 1.00$, $Q^2 = 0.88$) se obtuvieron valores muy altos de predicción. Las regiones del espectro asociado con las regiones de los ácidos grasos y ácidos orgánicos indicaron que fueron de particular relevancia para la clasificación del café Colombiano de acuerdo al gráfico de cargas (Figura 3.32 B y D), además en la clasificación de café Colombiano y Brasileño se encontraron que las variables de ácido acético y la cafeína también fueron relevantes.

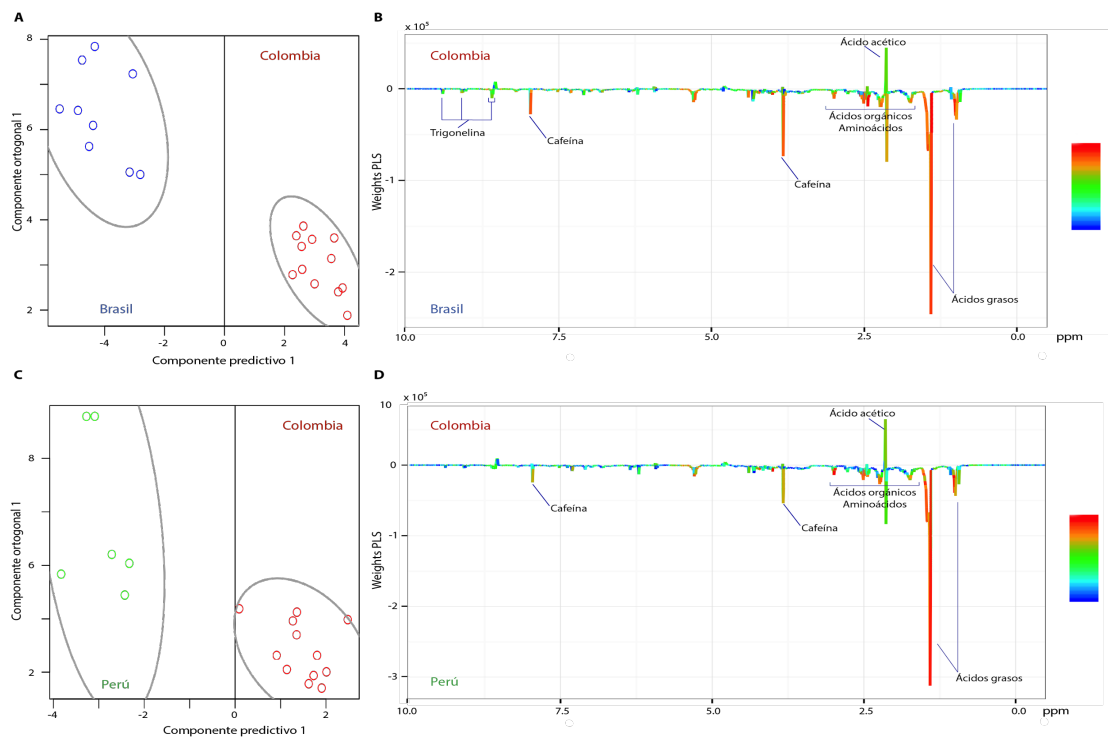


Figura 3.32. Gráfico de puntuaciones y cargas del análisis OPLS-DA de muestras de café obtenidos con huella metabólica ^1H -RMN para (A y B) Colombia vs Brasil, (C y D) Colombia vs Perú. Colombia (círculos rojos), Brasil (círculos azules) y Perú (círculos verdes).

4. CONCLUSIONES

En esta investigación se demuestra que es posible discriminar café Colombiano frente a cafés de otros orígenes aplicando metabolómica con el enfoque de huella metabolómica basada en ^1H -RMN. Las condiciones RMN experimentales implementadas aseguran una alta reproducibilidad de los espectros adquiridos, condición indispensable para trabajar con este enfoque.

El desarrollo e implementación de una serie de experimentos de RMN para obtener presaturación múltiple permitió trabajar con solvente sin sustitución isotópica de los protones con deuterio, logrando construir un sistema de bajo costo, robusto y totalmente automático para la adquisición de los espectros de 16 muestras de café verde o tostado en un tiempo de 3 horas. El análisis de los extractos de café obtenidos con la solución tampón de fosfato fueron de buena calidad pero la muestra presentó una rápida degradación (< 1 hora) dificultando la automatización del análisis. Este problema fue superado empleando metanol como solvente de extracción.

En los espectros ^1H -RMN de extractos metanólicos de café de las especies *C. arabica* y *C. canephora* var. *robusta* se podían observar claras diferencias en las intensidades de las señales de algunos metabolitos, principalmente en las regiones de las señales de cafeína, ácidos orgánicos y ácidos clorogénicos, así como una señal en 0.98 ppm que solo se encuentra presente en la especie *C. canephora*. En el caso de adulteración de la especie *C. arabica* por mezcla con la especie *C. canephora* también se observaron algunos cambios en el espectro, aumento en la intensidad y desplazamiento hacia campo alto de algunas señales a medida que aumentaba la proporción de *C. canephora*. Sin embargo, en cantidades menores de 20 % de *C.*

canephora resultaba difícil establecer a simple vista en el espectro si existía alguna adulteración.

La estrategia de control basada en el modelo PCA permitió detectar cambios en las condiciones ambientales o en el sistema de medida, al identificar muestras con comportamiento atípico con respecto al resto del grupo, las cuales fueron excluidas del conjunto de datos debido a que presentaban una pobre calidad del espectro, como mala supresión de las señales del solvente, fallas en el shim y corrección de la línea base. Además, el PCA permitió identificar agrupamiento con respecto al origen botánico, caso contrario con el origen geográfico.

El modelo de clasificación de origen botánico obtenido por PLS-DA para muestras de café verde presentó excelentes valores de predicción ($Q^2 = 1.0$) tanto en el conjunto de prueba como en el conjunto de validación externa. Con especies puras los resultados fueron excelentes. Los valores de predicción obtenidos por PLS-DA con muestras de café tostado (grado de tostado medio) fueron algo menores, $Q^2 > 0.89$, debido posiblemente a que existen muestras con contenido de *C. canephora* por debajo del límite predictivo de detección, 20 %. Mediante el modelo PLS ($R^2 = 0.9826$ y $Q^2 = 0.9759$) se puede cuantificar la cantidad de *C. canephora* en una mezcla.

El modelo de clasificación de origen geográfico obtenido por PLS-DA tanto para muestras de café verde como tostado, presentó valores muy buenos de predicción ($Q^2 > 0.74$ para el conjunto de prueba y $Q^2 > 0.87$ para el conjunto de validación externa). Por lo general, algunas muestras de café de países vecinos a Colombia se clasificaban de forma incorrecta como cafés Colombianos. La discriminación de los cafés Colombianos dependió en gran medida de metabolitos como sacarosa, cafeína, ácidos orgánicos y clorogénicos para el caso de café verde, y acetato, ácidos grasos y clorogénicos para el caso de café tostado.

En el modelo PLS-DA fue posible obtener valores cercanos de predicción para el conjunto de validación externa cuando se incluyeron muestras con grado de tostado, oscuro medio y oscuro, sin necesidad de modificar el conjunto de entrenamiento para la construcción del modelo, simplemente con la exclusión de una zona específica del espectro (*bins* 2.18 a 2.30 ppm correspondiente a la región de los ácidos orgánicos), se logró que el modelo fuera más robusto, es decir menos sensible al grado de tostado.

El perfil metabolómico del café con grado de tostado muy oscuro se ve fuertemente modificado, lo cual conlleva a clasificaciones incorrectas ($Q^2 < 0.20$ para el conjunto de validación externa). Por tanto, estas muestras fueron detectadas y excluidas antes de la determinación del origen botánico y geográfico mediante un modelo PLS.

Al comparar la capacidad discriminante de tres técnicas analíticas (RMN, GC-MS y GC-C-IRMS) para la clasificación de cafés tostado de países vecinos, Colombia, Brasil y Perú, se encontró que la metodología huella metabolómica ^1H -RMN, dado que permite la observación simultánea de un gran número de metabolitos, proporciona mejores resultados que las otras dos técnicas.

Las metodologías de extracción y análisis de metabolitos tanto en café verde como en café tostado combinadas con análisis quimiométrico pueden ser útiles para la industria del café como una herramienta rápida, robusta y precisa para la clasificación del café según el origen botánico (*C. arabica*, *C. canephora* o mezcla de las dos especies) y geográfico, siendo complementario a los análisis de calidad sensorial (catación), granulometría y humedad, para la certificación de calidad que debe cumplir el café Colombiano de exportación.

5. REFERENCIAS

Alessandrini, L., Romani, S., Pinnavaia, G., & Rosa, M.D. (2008). Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*, 625(1), 95–102.

Alonso-Salces, R.M., Serra, F., Reniero, F., & Heberger, K. (2009). Botanical and Geographical Characterization of Green Coffee (*Coffea arabica* and *Coffea canephora*): Chemometric Evaluation of Phenolic and Methylxanthine Contents. *J. Agric. Food Chem.*, 57, 4224–4235.

Alonso-Salces, R.M., Héberger, K., Holland, M.V., Moreno-Rojas, J.M., Mariani, C., Bellan, G., Reniero, F., & Guillou, C. (2010a). Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chemistry*, 118, 956–965.

Alonso-Salces, R.M., Moreno-Rojas, J.M., Holland, M.V., Reniero, F., Guillou, C., & Héberger, K. (2010b). Virgin Olive Oil Authentication by Multivariate Analyses of ^1H -NMR Fingerprints and $\delta^{13}\text{C}$ and $\delta^2\text{H}$ Data. *Journal of Agricultural and Food Chemistry*, 58, 5586–5596.

Altman, D.G., & Bland, J.M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 308(6943), 1552.

Alves, M.R., Casal, S., Oliveira, M.B., & Ferreira, M.A. (2003). Contribution of FA Profile Obtained by High-Resolution GC/Chemometric Techniques to the Authenticity of Green and Roasted Coffee Varieties. *JAOCs*, 80(6), 511-517.

Alves, R.C., Casal, S., Alves, M.R., & Oliveira, M.B. (2009). Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. *Food Chemistry*, 114, 295-299.

Amman, C., Meier, P., & Merbach, A.E. (1982). A Simple Multinuclear NMR Thermometer. *Journal of Magnetic Resonance*, 46, 319–321.

Anastasiadi, M., Zira, A., Magiatis, P., Haroutounian, S.A., Skaltsounis, A.L., & Mikros, E. (2009). ¹H-NMR-Based Metabonomics for the Classification of Greek Wines According to Variety, Region, and Vintage. Comparison with HPLC Data. *Journal of Agricultural and Food Chemistry*, 57, 11067–11074.

Anderson, K.A., & Smith B.W. (2002) Chemical Profiling To Differentiate Geographic Growing Origins of Coffee. *J. Agric. Food Chem.*, 50, 2068–2075.

Andrade, P.B., Leitão, R., Seabra, R.M., Oliveira, M.B., & Ferreira, M.A. (1998). 3,4-Dimethoxycinnamic acid levels as a tool for differentiation of *Coffea canephora* var. robusta and *Coffea arabica*. *Food Chemistry*, 61(4), 511-514.

Baker, M., & Rayens, W. Partial least squares for discrimination. *Journal of Chemometrics* (2003), 17(3), 166-173.

Bekedam, E.K., Loots, M.J., Schols, H.A., Van Boekel, M.A.J.S., & Smit, G. (2008). Roasting effects on formation mechanisms of coffee brew melanoidins. *J. Agric. Food Chem.*, 56, 7138–7145.

Bosco, M., Toffanin, R., De Palo, D., Zatti, L., & Segre, A. (1999). High-resolution ¹H NMR investigation of coffee. *J. Sci. Food Agric.*, 79, 869–878.

Briandet, R., Kemsley, E.K., & Wilson, R.H. (1996a). Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. *Journal of Science of Food and Agriculture*, 71, 359-366.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometrics*, 20, 341–351.

Briandet, R., Kemsley, E.K., & Wilson, R.H. (1996b). Discrimination of Arabica in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *Journal of Agricultural and Food Chemistry*, 44, 170-174.

Broadhurst, D.I.; & Kell, D.B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics.*, 2(4), 171-196.

Cagliani, L.R., Pellegrino, G., Giugno, G., & Consonni, R. (2013). Quantification of

Coffea arabica and *Coffea canephora* var. robusta in roasted and ground coffee blends. *Talanta*, 106, 169–173.

Capuano, E., & van Ruth, S.M. (2013). Analytical authentication of organic produce: An overview of markers. *Journal of the Science of Food and Agriculture*, 93, 12–28.

Carr, H.Y., & Purcell, E.M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review*, 94(3), 630-638.

Carrera, F., León-Camacho, M., Pablos, F., & González, A.G. (1998). Authentication of green coffee varieties according to their sterolic profile. *Anal. Chim. Acta*, 370, 131-139.

Casal, S., Oliveira, M.B., Alves, M.R., Ferreira, M.A. (2000). Discriminate Analysis of Roasted Coffee Varieties for Trigonelline, Nicotinic Acid, and Caffeine Content. *J. Agric. Food Chem.*, 48(8), 3420–3424.

Casal, S., Alves, M.R., Mendes, E.L., Oliveira, M.B.P.P., & Ferreira, M.A. (2003). Discrimination between Arabica and Robusta Coffee Species on the Basis of Their Amino Acid Enantiomers. *J. Agric. Food Chem.*, 51, 6495–6501.

Cevallos-Cevallos, J.M., Reyes-De-Corcuera, J.I., Etxeberria, E., Danyluka, M.D., & Rodrick, G.E. (2009). Metabolomic analysis in food science: a review. *Trends in Food Science & Technology*, 20, 557-566.

Charlton, A.J., Farrington, W.H., & Brereton, P. (2002). Application of ¹H NMR and Multivariate Statistics for Screening Complex Mixtures: Quality Control and Authenticity of Instant Coffee. *J. Agric. Food Chem.*, 50, 3098–3103.

Choi, M-Y., Choi, W., Park, J.H., Lim, J., & Kwon, S.W. (2010). Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. *Food Chemistry*, 121(4), 1260-1268.

Ciampa, A., Renzi, G., Taglienti, A., Sequi, P., & Valentini, M. (2010). Studies On Coffee Roasting Process By Means Of Nuclear Magnetic Resonance Spectroscopy. *J. Food Qual.*, 33, 199–211.

Clarke, R. In: *Coffee*, 1st edition, Clarke, R. J., & Macrae, R., eds. Essex, UK: Elsevier Applied Science Publishers; 1985, Vol. 1: Chemistry, p. 115.

Clifford, M.N. (2000). Chlorogenic acids and others cinnamates-nature, occurrence, dietary burden, absorption and metabolism. *Journal of Science and Food Agricultural*, 80, 1033–1043.

Clifford, M.N., Johnston, K.L., Knight, S., & Kuhnert, N. (2003). The characterisation by LC-MSⁿ of coffee seed caffeoylferuloylquinic acids. *J. Agric. Food Chem.*, 51, 2900–2911.

Consonni, R., Cagliani, L.R., & Cogliati, C. (2012). NMR based geographical characterization of roasted coffee. *Talanta*, 88, 420-426.

Costa Freitas, A.M., & Mosca, A.I. (1999). Coffee geographic origin- an aid to coffee differentiation. *Food Research International*, 32(8), 565-573.

Craig, H. (1957). Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochim. Cosmochim. Acta*, 12, 133–149.

Cruciani, G., Baroni, M., Clementi, S., Costantino, G., Riganelli, D., & Skagerberg, B. (1992). Predictive ability of regression models. Part I: Standard deviation of prediction errors (SDEP). *Journal of Chemometrics*, 6(6), 335–346.

D'Amelio, N., Fontanive, L., Uggeri, F., Suggi-Liverani, F., & Navarini, L. (2009). NMR Reinvestigation of the Caffeine–Chlorogenate Complex in Aqueous Solution and in Coffee Brews. *Food Biophysics*, 4:321–330.

Dias, R.C., Campanha, F.G., Vieira, L.G., Ferreira, L.P., Pot, D., Marraccini, P., & De Toledo Benassi, M. (2010). Evaluation of Kahweol and Cafestol in Coffee Tissues and Roasted Coffee by a New High-Performance Liquid Chromatography Methodology. *J. Agric. Food Chem.*, 58, 88–93.

Drivelos, S.A., & Georgiou, C.A. (2012). Multi-element and multi-isotope- ratio analysis to determine the geographical origin of foods in the European Union. *Trends in Analytical Chemistry*, 40, 38-51.

Duarte, G., & Farah, A. (2009). Chlorogenic acids and lactones in Brazilian commercial coffee. Proc. 22nd Int. Conf. Coffee Sci. ASIC/Prospero, 224–227. Trieste, Italy.

Duarte, G.S., Pereira, A.A., & Farah, A. (2010). Chlorogenic acids and other relevant compounds in Brazilian coffees processed by semi-dry and wet post-harvesting methods. *Food Chemistry*, 118, 851–855.

Dunn, W.B., & Ellis, D.I. (2005). Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry*, 24(4), 285-294.

Ebrahimi-Najafabadi, H., Leardi, R., Oliveri, P., Casolino, M.C., Jalali-Heravi, M., & Lanteri, S. (2012). Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. *Talanta*, 99, 175–179.

El-Abassy, R.M., Donfack, P., & Materny, A. (2011). Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis. *Food Chemistry*, 126, 1443-1448.

Esteban-Díez, I., González-Sáiz, J.M., & Pizarro, C. (2004). An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS. *Analytica Chimica Acta*, 514, 57-67.

Esteban-Díez, I., González-Sáiz, J.M., Sáenz-González, C., & Pizarro, C. (2007). Coffee varietal differentiation based on near infrared spectroscopy. *Talanta*, 71, 221-229.

Farah, A., de Paulis, T., Trugo, L.C., & Martin, P.R. (2005). Formation of chlorogenic acids lactones in roasted coffee. *Journal of Agricultural and Food Chemistry*, 53, 1105–1113.

Farah, A., de Paulis, T., Moreira, D.P., Trugo, L.C., & Martin, P.R. (2006a). Chlorogenic acids and lactones in regular and water-decaffeinated Arabica coffees. *Journal of Agricultural and Food Chemistry*, 54, 374–381.

Farah, A., & Donangelo, C.M. (2006b). Phenolic compounds in coffee. *Brazilian Journal Plant Physiology*, 18(1), 23–36.

Farah, A., Monteiro, M.C., Calado, V., Franca, A.S., & Trugo, L.C. (2006c). Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry*, 98, 373–380.

Federación Nacional de Cafeteros de Colombia (FNC). (2010a) Café de Colombia. Extraído el 15 de septiembre de 2014, de http://www.cafedecolombia.com/particulares/en/el_cafe_de_colombia/

Federación Nacional de Cafeteros de Colombia (FNC). (2010b) Garantía de Origen. Extraído el 15 de septiembre de 2014, de http://www.cafedecolombia.com/particulares/en/indicaciones_geograficas/Indicacion_Geografica_Protegida/

Federación Nacional de Cafeteros de Colombia (FNC). (2010c) Garantía de Origen. Extraído el 15 de septiembre de 2014, de http://www.cafedecolombia.com/particulares/en/indicaciones_geograficas/

Fiehn, O. (2002). Metabolomics-the link between genotypes and phenotypes, *Plant Mol. Biol.*, 48, 155–171.

Fischer, M., Reimann, S., Trovato, V., & Redgwell, R. J. (2001). Polysaccharides of green Arabica and Robusta coffee beans. *Carbohydr. Res.*, 330, 93–101.

Flament, I. (2002). Coffee Flavour Chemistry (John Wiley & Sons, Eds). Chichester, England. pp. 3-33, 37-52. ISBN: 978-0-471-72038-6.

Folstar, P., Van der Plas, H. C., Pilnik, W., & De Heusk J. G. (1977). Tocopherols in the unsaponifiable matter of coffee bean oil. *J. Agric. Food Chem.*, 25, 283–285.

Folstar, P. Lipids. In: Coffee, Clarke, R. J., & Macrae, R., eds. London: Elsevier Applied Science; 1985, Vol. 1: Chemistry, pp. 203–222.

Gallego, J.C. (2010). Proceso de calificación y sello de calidad en relación con el origen, caso: Café de Colombia. Consultoría realizada para la FAO y el ICA en el marco del estudio conjunto sobre los productos de calidad vinculada al origen (pp. 19-41). Documento electrónico.

Garret, R., Vaz, B.G., Hovell, A.M.C., Eberlin, M.N., & Rezende, C.M. (2012). Arabica and robusta coffees: identification of major polar compounds and quantification of blends by direct-infusion electrospray ionization-mass spectrometry. *J. Agric. Food Chem.*, 60, 4253–4258.

Garrett, R., Schmidt, E.M., Pereira, L.F.P., Kitzberger, C.S.G., Scholz, M.B.S., Eberlin, M.N., & Rezende, C.M. (2013). Discrimination of arabica coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics. *LWT - Food Science and Technology*, 50(2), 496-502.

Ginz, M., Balzer, H.H., Bradbury, A.G.W., & Maier, H.G. (2000). Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *Eur. Food Res. Technol.*, 211, 404-410.

González, A.G., Pablos, F., Martín, M.J., León-Camacho, L., & Valdenebro, M.S. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry*, 73, 93-101.

Goodacre, R., Broadhurst, D., Smilde, A.K., Kristal, B.S., Baker, J.D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D.B., Manetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjöström, M., Trygg, J., & Wulfert, F. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3:231-241.

Haiduc, A., Gancel, C., & Leloup, V. (2006). NIR-based determination of differences in green coffee chemical composition due to geographical origin. In 21st International Conference on Coffee Science (pp. 143-149), Montpellier, France.

Halouska, S., & Powers, R. (2006). Negative impact of noise on the principal component analysis of NMR data. *Journal of Magnetic Resonance*, 178(1), 88-95.

Holscher, W., Vitzthum, O.G., & Steinhart, H. (1990). Identification and sensorial evaluation of aroma-impact-compounds in roasted Colombian coffee. *Cafe, Cacao, The.*, 34, 205-212.

Hotelling, H. (1931). The generalization of Student's ratio. *Ann. Math. Stat.*, 2, 360-378.

Hovell, A.M.C., Pereira, E.J., Arruda, N.P., & Rezende, C. M. (2010). Evaluation of alignment methods and data pretreatments on the determination of the most important peaks for the discrimination of coffee varieties Arabica and Robusta using gas chromatography-mass spectroscopy. *Analytica Chimica Acta*, 678(2), 160-168.

Ibáñez, C., García-Cañas, V., Valdés, A., & Simó, C. (2013). Novel MS-based approaches and applications in food metabolomics. *Trends in Analytical Chemistry*, 52, 100–111.

International coffee organization (ICO). (2012). Annual review. Extraído el 14 de enero de 2015, de <http://www.ico.org/documents/cy2012-13/annual-review-2011-12e.pdf>

Jena, J.L., & Watling, R.J. (2013). Provenance establishment of coffee using solution ICP-MS and ICP-AES. *Food Chemistry*, 141(1), 98-104.

Jham, G.N., Winkler, J.K., Berhow, M.A., & Vaughn, S.F. (2007). γ -Tocopherol as a marker of Brazilian coffee (*Coffea arabica* L.) adulteration by corn. *Journal of Agricultural and Food Chemistry*, 55, 5995–5999.

Jove, M., Serrano, J.C.E., Ortega, N., Ayala, V., Angles, N., Reguant, J., Morello, J.R., Romero, M.P., Motilva, M.J., Prat, J., Pamplona, R., & Portero-Otín, M. Multicompartmental LC-Q-TOF-Based Metabonomics as an Exploratory Tool to Identify Novel Pathways Affected by Polyphenol-Rich Diets in Mice. *J. Proteome Res.* (2011), 10, 3501–3512.

Jumhawan, U., Putri, S.P., Yusianto, Marwani, E., Bamba, T., & Fukusaki, E. (2013). Selection of Discriminant Markers for Authentication of Asian Palm Civet Coffee (Kopi Luwak): A Metabolomics Approach. *J. Agric. Food Chem.*, 61, 7994–8001.

Kamm, K., Dionisi, F., Fay, L-B., Hischenhuber, C., Schmarr, H-G., & Engel, K-H. (2002). Rapid and Simultaneous Analysis of 16-O-Methylcafestol and Sterols as Markers for Assessment of Green Coffee Bean Authenticity by On-line LC–GC. *JAOCS*, 79(11), 1109-1113.

Keidel, A., von Stetten, D., Rodrigues, C., Máguas, C., & Hildebrandt, P. (2010). Discrimination of Green Arabica and Robusta Coffee Beans by Raman Spectroscopy. *J. Agric. Food Chem.*, 58, 11187–11192.

Kelly, S., Heaton, K., & Hoogewerff, J. (2005). Tracing the geographic origin of food: the application of multi-element and multi-isotope analysis, *Trends in Food Science and Technology*, 16, 555-567.

Kemsley, E.K., Ruault, S., & Wilson, R.H. (1995). Discrimination between *Coffea arabica* and *Coffea canephora* variant *robusta* beans using infrared spectroscopy. *Food Chemistry*, 54, 321-326.

Kölling-Speer, I., & Speer, K. Raw bean composition. In: Espresso Coffee, the Science of Quality. Illy, A., Viani, R., eds. Italy: Elsevier Academic Press; 2005, pp. 148–178.

Kwon, D-J., Jeong, H-J., Moon, H., Kim, H-N., Cho, J-H., Lee, J-E., Hong, K.S., & Hong, Y-S. (2015). Assessment of green coffee bean metabolites dependent on coffee quality using a ¹H NMR-based metabolomics approach. *Food Research International*, 67, 175–182.

Lee, J-E., Lee, B-J., Chung, J-O., Hwang, J-A., Lee, S-J., Lee, C-H., & Hong, Y-S. (2010). Geographical and Climatic Dependencies of Green Tea (*Camellia sinensis*) Metabolites: A ¹H-NMR-Based Metabolomics Study. *Journal of Agricultural and Food Chemistry*, 58, 10582–10589.

Link, J.V., Guimarães Lemes, A.L., Marquetti, I., dos Santos Scholz, M.B., & Bona, E. (2014). Geographical and genotypic classification of arabica coffee using Fourier transform infrared spectroscopy and radial-basis function networks. *Chemometrics and Intelligent Laboratory Systems*, 135, 150–156.

Liu, H-C., You, C-F., Chen, C-Y., Liu, Y-C., & Chung, M-T. (2014). Geographic determination of coffee beans using multi-element analysis and isotope ratios of boron and strontium. *Food Chemistry*, 142, 439–445.

Luykx, D.M.A.M., & van Ruth, S.M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, 107, 897–911.

Macrae, R. (1985). Nitrogenous compounds. In Clarke, R.J., & Macrae R. (Eds.), Coffee chemistry (pp. 115–152). Elsevier, Applied Science: Amsterdam.

Martín, M.J., Pablos, F., & González, A.G. (1998a). Discrimination between arabica and robusta green coffee varieties according to their chemical composition. *Talanta*, 46(6), 1259-1264.

Martín, M.J., Pablos, F., & González, A.G. (1998b). Characterization of green coffee varieties according to their metal content. *Analytica Chimica Acta*, 358, 177-183.

Martín, M.J., Pablos, F., & González, A.G. (1999). Characterization of *arabica* and *robusta* roasted coffee varieties and mixture resolution according to their metal content. *Food Chemistry*, 66(3), 365-370.

Martín, M.J., Pablos, F., González, A.G., Valdenebro, M.S., & León-Camacho, M. (2001). Fatty acid profiles as discriminant parameters for coffee varieties differentiation. *Talanta*, 54(2), 291-297.

Max, K. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt and Tony Cooper (2013). Caret: Classification and Regression Training. R package version 5.17-7. <http://CRAN.R-project.org/package=caret>

McKay, R.T. (2011). How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn. Reson. Part A*, 38A(5), 197-220.

McLafferty, F.W., Staufner, D.B. (1989). The Wiley/NBS Registry of Mass Spectral Data Wiley, New York.

Meiboom, S., & Gill, D. (1958). Modified SpinEcho Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.*, 29, 688-691.

Mendonça, J.C.F., Franca, A.S., Oliveira, L.S., & Nunes, M. (2008). Chemical characterisation of non-defective and defective green arabica and robusta coffees by electrospray ionization-mass spectrometry (ESI-MS). *Food Chemistry*, 111, 490-497.

Mevik, B-H., Wehrens, R., & Liland, K.H. (2013). pls: Partial Least Squares and Principal Component regression. R package version 2.4-3. <http://CRAN.R-project.org/package=pls>

Moreira, R.F.A., Trugo, L.C., & Maria, C.A.B. (2000). Compostos voláteis do café torrado. Parte II. Compostos alifáticos, alicíclicos e aromáticos. *Química Nova*, 23, 195-201.

Murkovic, M., & Derler, K. (2006). Analysis of amino acids and carbohydrates in green coffee *Journal of Biochemical and Biophysical Methods*, 69, 25-32.

Nikolova-Damyanova, B., Velikova, R., & Jham, G.N. (1998). Lipid classes, fatty acid composition and tri-acylglycerol molecular species in crude coffee beans harvested in Brazil. *Food Res. Int.*, 31, 479–486.

Ogawa, A., Kamiya, C., & Iida, Y. (1989). Contents of tocopherols in coffee beans, coffee infusions and instant coffee. *Nippon Shokuhin Kogyo Gakkaishi*, 36, 490–494.

Oliveira, R.C.S., Oliveira, L.S., Franca, A.S., & Augusti, R. (2009). Evaluation of the potential of SPME-GC-MS and chemometrics to detect adulteration of ground roasted coffee with roasted barley. *Journal of Food Composition and Analysis*, 22, 257–261.

Özdestan, Ö., van Ruth, S.M., Alewijn, M., Koot, A., Romano, A., Cappellin, L., & Biasioli, F. (2013). Differentiation of specialty coffees by proton transfer reaction-mass spectrometry. *Food Research International* 53 (2013) 433–439.

Pacetti, D., Boselli, E., Balzano, M., & Frega, N.G. (2012). Authentication of Italian Espresso coffee blends through the GC peak ratio between kahweol and 16-O-methylcafestol. *Food Chemistry*, 135, 1569–1574.

Pauli, E.D., Barbieri, F., Garcia, P.S., Madeira, T.B., Acquaro Junior, V.R., Ieda Spacino Scarminio, Paulinetti da Camara, C.A., & Nixdorf, S.L. (2014). Detection of ground roasted coffee adulteration with roasted soybean and wheat. *Food Research International*, 61, 112–119.

Pérez-Enciso, M., & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum. Genet.*, 112, 581–592.

Pizarro, C., Esteban-Díez, I., & González-Sáiz, J.M. (2007). Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, 585(2), 266–276.

Puertas, G.I. (2006). La humedad controlada del grano preserva la calidad del café. *Avances Técnicos Cenicafé* No. 352. 2 p.

Puertas, G.I. (2010). Fundamentos del proceso de fermentación en el beneficio del café. *Avances Técnicos Cenicafé* No. 402. 12 p.

Puertas, G.I. (2012). Factores, procesos y controles en la fermentación del café. *Avances Técnicos Cenicafé* No. 422. 12 p.

R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN t3-900051-07-0, URL <http://www.R-project.org>.

Rannar, R., Lindgren, F., Geladi, P., & Wold, S. (1994). A PLS kernel algorithm for data sets with many variables and fewer objects. Part I. Theory and Algorithm. *Journal of Chemometrics*, 8, 111–125.

Reid, L., O'Donnel, C.P., & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science and Technology*, 17, 344–353.

Reis, N., Franca, A.S. & Oliveira, L.S. (2013). Performance of diffuse reflectance infrared Fourier transform spectroscopy and chemometrics for detection of multiple adulterants in roasted and ground coffee. *LWT - Food Science and Technology*, 53, 395-401.

Risticevic, S., Carasek, E., & Pawliszyn, J. (2008). Headspace solid-phase microextraction-gas chromatographic-time-of-flight mass spectrometric methodology for geographical origin verification of coffee. *Analytica Chimica Acta*, 617 (1-2), 72-84.

Rodrigues, C.I., Maia, R., Miranda, M., Ribeirinho, M., Nogueira, J.M.F., & Máguas, C. (2009). Stable isotope analysis for green coffee bean: A possible method for geographic origin discrimination. *Journal of Food Composition and Analysis*, 22(5), 463-471.

Rodrigues, C., Brunner, M., Steiman, S., Bowen, J.G., Nogueira, J.M.F., Gautz, L., Prohaska, T., & Máguas, C. (2011a). Isotopes as Tracers of the Hawaiian Coffee-Producing Regions. *J. Agric. Food Chem.*, 59(18), 10239-10246.

Rodrigues, C., Máguas, C., & Prohaska, T. (2011b). Strontium and oxygen isotope fingerprinting of green coffee beans and its potential to proof authenticity of coffee. *Eur. Food Res. and Technol.*, 232(2), 361-373.

Romano, R., Santini, A., Grottaglie, L.L., Manzo, N., Visconti, A., & Ritieni, A. (2014). Identification markers based on fatty acid composition to differentiate between roasted *Arabica* and *Canephora* (Robusta) coffee varieties in mixtures. *Journal of Food Composition and Analysis*, 35, 1–9.

Rubayiza, A.B., & Meurens, M. (2005). Chemical Discrimination of Arabica and Robusta Coffees by Fourier Transform Raman Spectroscopy. *J. Agric. Food Chem.*, 53(12), 4654-4659.

Santato, A., Bertoldi, D., Perini, M., Camin, F., & Larcher, R. (2012). Using elemental profiles and stable isotopes to trace the origin of green coffee beans on the global market. *J. Mass Spectrom.*, 47, 1132–1140.

Santos, J.R., Sarraguça, M.C., Rangel, A.O., & Lopes, J.A. (2012). Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach. *Food Chemistry*, 135(3), 1828-1835.

Sarraguça, M.C., Santos, J.R., Rangel, A.O., & Lopes, J.A. (2013). Authenticity Control of Roasted Coffee Brands Using Near-Infrared Spectroscopy. *Food Analytical Methods*, 6(3), 892-899.

Serra, F., Guillou, C.G., Reniero, F., Ballarin, L., Cantagallo, M.I., Wieser, M. Iyer, S.S., Héberger, K., & Vanhaecke, F. (2005). Determination of the geographical origin of green coffee by principal component analysis of carbon, nitrogen and boron stable isotope ratios. *Rapid Commun. Mass Spectrom.*, 19, 2111-2115.

Stadler, R. H., Varga, N., Hau, J., Vera, F. A., & Welti, D. H. (2002). Alkylpyridiniums. 1. Formation in model systems via thermal degradation of trigonelline. *J. Agric. Food Chem.*, 50, 1192–1199.

Smit, S., van Breemen, M.J., Hoefsloot, H.C.J., Smilde, A.K., Aerts, J.M.F.G., & de Koster, C.G. (2007). Assessing the statistical validity of proteomics based biomarkers. *Analytica Chimica Acta*, 592(2), 210–217.

- Son, H-S., Hwang, G-S., Ahn, H-J., Park, W-M., Lee, C-H., & Hong, Y-S. (2009). Characterization of wines from grape varieties through multivariate statistical analysis of ^1H -NMR spectroscopic data. *Food Research International*, 42, 1483–1491.
- Souto, U.T.C.P., Pontes, M.J.C., Silva, E.C., Galvão, R.K.H., Araújo, M.C.U., Sanches, F.A.C. Cunha, F.A.S., & Oliveira, M.S.R. (2010). UV–Vis spectrometric classification of coffees by SPA–LDA. *Food Chemistry*, 119, 368.
- Speer, K., & Kölling-Speer, I. (2006). The lipid fraction of the coffee bean. *Braz. J. Plant Physiol.*, 18, 201–216.
- Spraul, M., Schütz, B., Rinke, P., Koswig, S., Humpfer, E., Schäfer, H., Mörtter, M., Fang, F., Marx, U. C., & Minoja, A. (2009). NMR-Based Multi Parametric Quality Control of Fruit Juices: SGF Profiling. *Nutrients*, 1, 148-155.
- Staub, C. (1995). Roast Color Classification System Developed by Agron-SCAAA. 17th Specialty Coffee Association of America ASIC, Long Beach, California.
- Suchánek, M., Filipova, H., Volka, K., Delgadillo, U., & Davies, A. N. (1996). Qualitative Analysis of Green Coffee by Infrared Spectroscopy. *Fresenius J. Anal. Chem.*, 354, 327-332.
- Szymańska, E., Saccenti, E., Smilde, J.A., & Westerhuis, A.K. (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, 8(1), 3-16.
- Tavares, L.A., & Ferreira, A.G. (2006). Análises Quali- E Quantitativa De Cafés Comerciais Via Ressonância Magnética Nuclear. *Quim. Nova*, 29(5), 911-915.
- Teixeira, A. A., Brando, C. H. J., Thomaziello, R. A., & Teixiera, R. (1995). In Illy A., & Viani, R. (Eds.), Espresso coffee: The science of quality (pp. 53, 91–95, 197-198). Italy: Elsevier Academic Press.
- Taguchi, H., Sakaguchi, M., & Shimabayashi, Y. (1985). Trigonelline content in coffee beans and the thermal conversion of trigonelline into nicotinic acid during the roasting of coffee beans. *Agric. Biol. Chem.*, 49, 3467–3471.

Toci, A. T., Farah, A., Trugo, L. C. (2006). Efeito do processo de descafeinação com diclorometano sobre a composição química dos cafês Arábica e Robusta antes e após a torração. *Química Nova*, 29, 965–971.

Toci, A. T., & Farah, A. (2008). Volatile compounds as potential defective coffee seeds' markers. *Food Chemistry*, 108, 1133–1141.

Tornincasa, P., Furlan, M., Pallavicini, & Graziosi, G. (2010). Coffee species and varietal identification. In Nimis, P.L., & Vignes Lebbe R. (eds.). *Tools for Identifying Biodiversity: Progress and Problems* (pp. 307-313). ISBN 978-88-8303-295-0. Trieste, Italia.

Trugo, L. C. Carbohydrates. In: *Coffee*, 1st edition, Clarke R. J., Macrae, R., eds. Essex: Elsevier Applied Science Publishers; 1985, Vol 1: Chemistry, p. 83.

Trugo, L.C., & Macrae R.A. (1984). study of the effect of roasting on the chlorogenic acid composition of coffee using HPLC. *Food Chemistry*, 15, 219–227.

Trugo, L.C. (2003). Coffee Analysis. In: *Encyclopedia of Food Science and Nutrition*, 2nd edition, Caballero, B., Trugo, L.C., Finglas, P.M., eds. Oxford, UK: Oxford Academic Press, Vol. 2, p. 498.

Trygg, J., & Wold, S. Orthogonal projections to latent structures (O-PLS). (2002). *Journal of Chemometrics*, 16(3), 119–128.

Tynkkynen, T., Tiainen, M., Soininen, P., & Laatikainen, R. (2009). From proton nuclear magnetic resonance spectra to pH. Assessment of ¹H NMR pH indicator compound set for deuterium oxide solutions. *Analytica Chimica Acta*, 648, 105–112.

United States Department of Agriculture (USDA). (2014) Coffee: World Markets and Trade. Extraído el 14 de enero de 2015 de <http://apps.fas.usda.gov/psdonline/circulars/coffee.pdf>

Valdenebro, M.S., León-Camacho, M., Pablos, F., González, A.G., & Martín, M.J. (1999). Determination of the arabica/robusta composition of roasted coffee according to their sterolic content. *Analyst*, 124, 999-1002.

Wang N., & Lim, L-T. (2012). Fourier Transform Infrared and Physicochemical Analyses of Roasted Coffee. *J. Agric. Food Chem*, 60, 5446-5453.

Weckerle, B., Richling, E., Heinrich, S., & Schreier, P. (2002). Origin assessment of green coffee (*Coffea arabica*) by multi-element stable isotope analysis of caffeine. *Anal. Bioanal. Chem.*, 374, 886.

Wei, F., Furihata, K., Hu, F., Miyakawa, T., & Tanokura, M. (2010). Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy. *Magn. Reson. Chem.*, 48, 857–865.

Wei, F., Furihata, K., Hu, F., Miyakawa, T., & Tanokura, M. (2011). Two-Dimensional ^1H - ^{13}C Nuclear Magnetic Resonance (NMR)-Based Comprehensive Analysis of Roasted Coffee Bean Extract. *J. Agric. Food Chem.*, 59, 9065–9073.

Wei, F., Furihata, K., Koda, M., Hu, F., Kato, R., Miyakawa, T., & Tanokura, M. (2012a). ^{13}C NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin. *J. Agric. Food Chem.*, 60(40), 10118–10125.

Wei, F., Furihata, K., Koda, M., Hu, F., Miyakawa, T., & Tanokura, M. (2012b). Roasting process of coffee beans as studied by nuclear magnetic resonance: time course of changes in composition. *J. Agric. Food Chem.* 60 (4), 1005–1012.

Wermelinger, T., D'Ambrosio, L., Klopprogge, B., & Yeretzyan, C. (2011). Quantification of the robusta fraction in a coffee blend via Raman spectroscopy: Proof of principle. *Journal of the Science of Food and Agriculture*, 59, 9074–9907.

Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., Smilde, A.K., Velzen, E.J.J., Duijnhoven, J.P.M., & Dorsten, F.A. (2008). Assessment of PLS-DA cross validation. *Metabolomics*, 4(1): 81–89.

Wishart, D.S. (2008). Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology*, 19(9), 482–493.

Wold, S., Sjöström, M., & Eriksson, L. (2001a). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.

Wold, S., Trygg, J., Berglund, A., & Antti, H. (2001b). Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 58, 131–150.

Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics*, 1, 92–107.

Zambonin, C.G., Balest, L., De Benedetto, G.E., & Palmisano, F. (2005). Solid-phase microextraction-gas chromatography mass spectrometry and multivariate analysis for the characterization of roasted coffees. *Talanta*, 66, 261-265.

6. ANEXOS

ANEXO 6.1

MUESTRAS UTILIZADAS EN EL ESTUDIO

Tabla A6.1a. Resumen de las muestras de semillas verdes de C. arabica y C. canephora var. robusta usadas en este estudio ordenadas por origen geográfico. Los números por fuera del paréntesis indican el número de muestras y por dentro del paréntesis indican el número de espectros. Las repeticiones de las muestras se realizaron al azar.

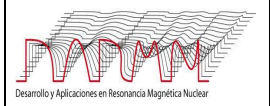
Continente	País	100% Arabica	100% Robusta
Sur América	Brasil	7(9)	0
	Colombia	46(48)	0
	Ecuador	1(1)	0
	Perú	3(4)	0
Norte América	Hawái	1(2)	0
Centro América	Costa Rica	1(1)	0
	Republica Dominicana	4(4)	0
	El Salvador	1(1)	0
	Guatemala	6(8)	1(1)
	Honduras	1(1)	0
	México	1(1)	1(1)
	Nicaragua	1(1)	0
África	Uganda	1(2)	3(6)
	Togo	0	1(1)
	Tanzania	1(2)	1(1)
	Costa de Marfil	0	1(1)
	Cameron	0	1(1)
Asia	China	4(4)	1(1)
	India	0	3(4)
	Indonesia	0	1(1)
	Vietnam	0	2(3)
Total	95(110)	79(89)	16(21)

Tabla A6.1b. Resumen de las muestras de semillas tostadas de C. arabica y C. canephora var. robusta usadas en este estudio ordenadas por origen geográfico. Los números por fuera del paréntesis indican el número de muestras y por dentro del paréntesis indican el número de espectros. Las repeticiones de las muestras se realizaron al azar.

Continente	País	100% Arabica	100% Robusta
Sur América	Brasil	26(33)	4(5)
	Colombia	101(115)	0
	Ecuador	3(4)	0
	Perú	17(23)	0
Norte América	Hawái	2(3)	0
Centro América	Costa Rica	7(9)	0
	Republica Dominicana	4(4)	1(1)
	El Salvador	3(3)	0
	Guatemala	16(16)	2(2)
	Honduras	8(8)	0
	Jamaica	1(1)	0
	México	6(6)	1(2)
	Nicaragua	4(4)	0
	Panamá	7(9)	0
África	Uganda	1(2)	4(6)
	Togo	0	1(1)
	Tanzania	2(2)	1(1)
	Etiopia	3(3)	0
	Costa de Marfil	0	1(2)
	Cameron	0	1(1)
Asia	China	4(5)	1(2)
	India	0	6(9)
	Indonesia	1(1)	3(3)
	Vietnam	0	7(9)
Total	249(295)	216(251)	33(44)

Mezcla de café Arabica con café Robusta	Contenido de café Robusta (%)	Número de espectros
	5	(7)
	10	(13)
	20	(15)
	30	(14)
	50	(16)
	70	(14)
	90	(16)
Total		(95)

ANEXO 6.2

 Desarrollo y Aplicaciones en Resonancia Magnética Nuclear	PROTOCOLO PARA LA PREPARACIÓN DE SOLUCIÓN TAMPÓN DE FOSFATO	Código: COFFEE-v.01 Fecha de actualización: 07/05/2012
---	--	--

1. REACTIVOS

- $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$
- $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$
- Agua grado Milli-Q

2. MATERIALES Y EQUIPOS

- Espátula
- Balones volumétricos de 250 mL
- Balanza analítica electrónica PA214C Serie PIONNER

3. PREPARACIÓN DE LA SOLUCIÓN TAMPÓN DE FOSFATO

3.1 Preparación de la solución Na_2HPO_4 , 100mM (Solución A)

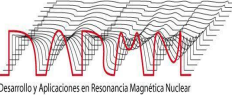
Pesar 2.2 g de $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ y disolver en agua. Llevar a un volumen final de 250 mL con agua desionizada.

3.2 Preparación de la solución NaH_2PO_4 , 100mM (Solución B)

Pesar 3.0 g de $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ y disolver en agua. Llevar a un volumen final de 250 mL con agua desionizada.

3.3 Preparación de la solución tampón $[\text{Na}_2\text{HPO}_4]/[\text{NaH}_2\text{PO}_4]$, 50.00mM/50.00mM

Mezclar las soluciones A y B en partes iguales.

 Desarrollo y Aplicaciones en Resonancia Magnética Nuclear	PROTOCOLO PARA LA PREPARACIÓN DE MUESTRAS DE CAFÉ VERDE Y TOSTADO EN SOLUCIÓN TAMPÓN PARA ANÁLISIS POR RMN	Código: COFFEE-v.01 Fecha de actualización: 07/05/2012
--	---	--

1. PRINCIPIO

La extracción de los metabolitos en muestras de café verde y tostado molido se realiza con solución tampón a pH 7.0 y agitación. El extracto es purificado por centrifugación y posteriormente se realiza el análisis por RMN.

2. REACTIVOS

- Solución tampón de fosfato a pH 7
- Oxido de deuterio, 99.9 atom %D
- Tetrametilsilano (TMS), 99.8 % (GC)

3. MATERIALES Y EQUIPOS

- Tubos para microcentrifuga de 1.5 mL
- Micropipeta, 100 – 1000 μ L y 0 – 100 μ L
- Puntas para micropipeta de 100 – 1000 μ L y 0 – 100 μ L
- Espátula
- Balanza analítica electrónica PA214C Serie PIONNER
- Vortex Genie 2 Scientific Industries. Model No. G-560E
- Microcentrifuga Eppendorf 5415R
- Tubos de RMN (Wilmad Labglass Inc., Vineland, NJ, USA)
- NMR Avance II 400MHz UltraShieldTM Bruker spectrometer

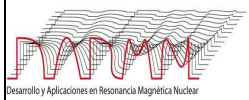
4. ALMACENAMIENTO DE LAS MUESTRAS DE CAFÉ

Las muestras de café verde y tostado molido son recibidas en el Laboratorio de RMN de la Universidad del Valle, cada muestra es etiquetada y guardada en nevera a 4 °C.

5. EXTRACCIÓN DE METABOLITOS EN MUESTRAS DE CAFÉ VERDE Y TOSTADO

200 ± 0.1 mg de café verde o tostado finamente molido se extrae con 1.5 mL de solución tampón de fosfato pH 7, luego se mezcla con un vortex durante 2 minutos y se centrifuga a 18 °C y 8608 g durante 10 minutos. Se repite el proceso de centrifugación con el sobrenadante en un nuevo tubo eppendorf y se transfiere 450 µL del sobrenadante al tubo de RMN junto con 50 µL de óxido de deuterio para el anclaje (lock) con TMS (referencia del desplazamiento químico). Cada serie de 15 muestras extraídas se realiza una repetición de una de ellas seleccionada al azar.

ANEXO 6.3

 Desarrollo y Aplicaciones en Resonancia Magnética Nuclear	PROTOCOLO PARA LA PREPARACIÓN DE MUESTRAS DE CAFÉ VERDE Y TOSTADO EN METANOL PARA ANÁLISIS POR RMN	Código: COFFEE-v.01 Fecha de actualización: 07/05/2012
---	---	--

1. PRINCIPIO

La extracción de los metabolitos en muestras de café verde y tostado molido se realiza con metanol sin sustitución isotópica y agitación. El extracto es purificado por centrifugación y posteriormente se realiza el análisis por RMN.

2. REACTIVOS

- Metanol 99.8 %, Grado HPLC
- Metanol-d₄, 99.8 atom %D
- Tetrametilsilano (TMS), 99.8 % (GC)

3. MATERIALES Y EQUIPOS

- Tubos para microcentrífuga de 1.5 mL
- Micropipeta, 100 – 1000 µL y 0 – 100 µL
- Puntas para micropipeta de 100 – 1000 µL y 0 – 100 µL
- Espátula
- Balanza analítica electrónica PA214C Serie PIONNER
- Vortex Genie 2 Scientific Industries. Model No. G-560E
- Microcentrífuga Eppendorf 5415R
- Tubos de RMN (Wilmad Labglass Inc., Vineland, NJ, USA)
- NMR Avance II 400MHz Bruker UltraShield™ spectrometer

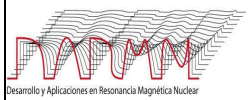
4. ALMACENAMIENTO DE LAS MUESTRAS DE CAFÉ

Las muestras de café verde y tostado molido son recibidas en el Laboratorio de RMN de la Universidad del Valle, cada muestra es etiquetada y guardada en nevera a 4 °C.

5. EXTRACCIÓN DE METABOLITOS EN MUESTRAS DE CAFÉ VERDE Y TOSTADO

200 ± 0.1 mg de café verde o tostado finamente molido se extrae con 1.0 mL de metanol grado cromatografico, luego se mezcla en un vortex durante 2 minutos y se centrifuga a 18 °C y 8608 g durante 10 minutos. 450 µL del sobrenadante es transferido al tubo de RMN junto con 50 µL de metanol deuterado para el anclaje (lock) con TMS (referencia del desplazamiento químico). Cada serie de 15 muestras extraídas se realiza una repetición de una de ellas seleccionada al azar.

ANEXO 6.4

	PROTOCOLO PARA LA PREPARACIÓN DE MUESTRAS DE CAFÉ TOSTADO PARA ANÁLISIS POR GC-MS	Código: COFFEE-v.01 Fecha de actualización: 26/05/2013
---	--	--

1. PRINCIPIO

La extracción de los metabolitos en muestras de café tostado molido se realiza con diclorometano y luego es filtrado, para posterior análisis por GC-MS.

2. REACTIVOS

- Diclorometano, puro, absoluto, sobre tamiz molecular ($H_2O \leq 0.005 \%$), $\geq 99.5 \%$ (GC). 66749 *SIGMA-ALDRICH*
- 1-Decanol, $\geq 99 \%$. 239763 *ALDRICH*

3. MATERIALES Y EQUIPOS

- Tubos para microcentrifuga de 2.0 mL
- Micropipeta, 100 – 1000 μ L
- Puntas para micropipeta de 100 – 1000 μ L
- Espátula
- Balanza analítica electrónica Mettler Toledo XP26
- Vortex Genie 2 Scientific Industries. Model No. G-560E
- Filtro de jeringa Whatman® ReZist®, diam. 30 mm, tamaño de poro 0.45 μ m, membrana de PTFE, no estéril. WHA10463513 *SIGMA*
- Viales de vidrio de 2.0 mL con tapa, septa PTFE/rojo silicona. 5182-0553 *AGILENT*
- Insertos de vidrio de 250 μ L. 5183-2085 *AGILENT*

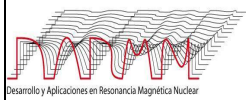
4. ALMACENAMIENTO DE LAS MUESTRAS DE CAFÉ

Las muestras de café tostado molido son recibidas en el Instituto de Ciencias Forenses de la Universidad de Lausanne - Suiza, cada muestra es guardada en nevera a 4 °C.

5. EXTRACCIÓN DE METABOLITOS EN MUESTRAS DE CAFÉ VERDE Y TOSTADO

200 ± 0.1 mg de café tostado finamente molido se extrae con 1 mL de diclorometano (estandar interno, 1-decanol 0.2 mg/mL), luego se mezcla en un vortex durante 2 min. La muestra es filtrada y transferida a un vial. La extracción de cada muestra se realiza por duplicado.

ANEXO 6.5

 Desarrollo y Aplicaciones en Resonancia Magnética Nuclear	PROTOCOLO PARA LA PREPARACIÓN DE MUESTRAS DE CAFÉ TOSTADO PARA ANÁLISIS POR GC-C-IRMS	Código: COFFEE-v.01 Fecha de actualización: 26/05/2013
--	--	--

1. PRINCIPIO

La extracción de los metabolitos en muestras de café tostado molido se realiza con agua caliente y posteriormente una extracción líquido-líquido con cloroformo. La fase orgánica es secada con Na₂SO₄ anhidro y filtrada, para posterior análisis por GC-C-IRMS.

2. REACTIVOS

- Agua Milli-Q
- Cloroformo ≥ 99.5 %(GC). 25693 SIGMA-ALDRICH
- Sulfato de sodio, anhidro, ≥ 99.0 %(T). 71962 FLUKA
- Ácido tetradecanoico metil ester, material de referencia certificado, TraceCERT[®]. 55791 FLUKA

3. MATERIALES Y EQUIPOS

- Tubos para microcentrífuga de 2.0 mL
- Micropipeta, 100 – 1000 μ L y 0 – 100 μ L
- Puntas para micropipeta de 100 – 1000 μ L y 0 – 100 μ L
- Espátula
- Balanza analítica electrónica Mettler Toledo XP26
- Vortex Genie 2 Scientific Industries. Model No. G-560E
- Termometro
- Filtro de jeringa Whatman[®] ReZist[®], diam. 30 mm, tamaño de poro 0.45 μ m, membrana de PTFE, no estéril. WHA10463513 SIGMA

- Jeringas de 5.0 mL BD Plastipak
- Viales de vidrio de 2.0 mL con tapa, septa PTFE/rojo silicona. 5182-0553AGILENT
- Insertos de vidrio de 250 μ L. 5183-2085AGILENT

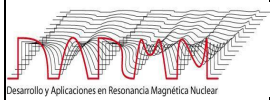
4. ALMACENAMIENTO DE LAS MUESTRAS DE CAFÉ

Las muestras de café tostado molido son recibidas en el Instituto de Ciencias Forenses de la Universidad de Lausanne - Suiza, cada muestra es guardada en nevera a 4 °C.

5. EXTRACCIÓN DE METABOLITOS EN MUESTRAS DE CAFÉ TOSTADO

80 $\pm$ 0.1 mg de café tostado finamente molido se extrae con 1 mL de agua caliente (90 °C), luego se mezcla en un vortex durante 10 min. La solución filtrada es sujeta a extracción líquido-líquido con 1 mL de cloroformo y mezclada en un vortex durante 10 min. La fase orgánica es separada y secada a través de una columna que contiene Na₂SO₄ anhidro, luego es filtrada. 180 μ L del extracto y 20 μ L de la solución del estándar interno, ácido tetradecanoico metil ester (40 μ g/mL) son transferidos a un vial. La extracción de cada muestra se realiza por duplicado.

ANEXO 6.6

	PROTOCOLO PARA LA OBTENCIÓN DE ESPECTROS ¹H-RMN DE MUESTRAS DE CAFÉ VERDE Y TOSTADO	Código: COFFEE-v.01 Fecha de actualización: 07/05/2012
---	---	--

El análisis ¹H-RMN de los extractos metanolicos de café verde y tostado se realizan después de realizar el proceso de optimización y calibración del espectrometro, para obtener el espectro se debe realizar una secuencia de tres experimentos 1D con el fin de optimizar los parámetros para la supresión del solvente de cada muestra. Estos experimentos se realizan de manera automática previa configuración en el iconnmr (automuestreador) de la secuencias de muestras a analizar (16 muestras por ejecución).

Experimento 1: El primer experimento consiste en un sencillo experimento (zg30) de un impulso de excitación corto (0,1 ms) y un tiempo de observación de 4s (AQ). Después de 4 observaciones vacías (*dummy scans*), se registraron y sumaron 8 observaciones (FID, free induction decay) para formar un vector complejo de 64k puntos con una ventana espectral de 20.5181 ppm y una ganancia del receptor RG = 90.5. Una vez terminada la transformación de Fourier (sin apodización), se aplica una rutina de búsqueda de picos automática para determinar las frecuencias de las señales de metanol y agua (3.3 y 4.7 ppm). Estos dos valores se utilizan para construir un impulso modulado en coseno de 1 s (25 Hz) con 50.000 puntos complejos para saturar selectivamente bandas en el siguiente experimento.

Parámetros de adquisición

Secuencia de pulso: zg30

AQ_mod: DQD

AUNM: au_prof1d_extract_zg30

Current pulse program

Acquisition mode

Acquisition AU program

AUNMP: proc_prof1d_extract pilot.lims	Processing AU program
TD: 65536	Size of fid
NS: 8	Number of scans
DS: 4	Number of dummy scans
TD0: 1	Loop count for 'td0'
SW [ppm]: 20.5581	Spectral width
SWH [Hz]: 8223.685	Spectral width
AQ [s]: 3.9846	Acquisition time
RG: 90.5	Receiver gain
DW [us]: 60.800	Dwell time
D1 [s]: 1.0	Delays
D8 [s]: 0.020	Delays
D16 [s]: 0.0002	Delays
PL1[dB]: -0.40	Power level
PL9[dB]: 59.14	Power level
O1[Hz]: 1882.49	Transmitter frequency offset

Experimento 2: El segundo experimento (zgps) consiste en un experimento 1D con supresión de las señales de agua y metanol durante el retraso de relajación, esto es RD - P (90 °) - AQ. Un campo de saturación de banda selectiva 6s es obtenido por la unión de 6 impulso modulado en coseno construido con las frecuencias de disolvente medidos en el primer experimento. Después de dos *dummy scans* (DS = 2), un vector de 64k puntos de datos complejos (AQ = 4s) se registra mediante la adición de 4 FID (NS = 4) con una anchura de barrido de 20.5181 ppm y una ganancia del receptor de 4. El espectro resultante se utiliza para refinar las frecuencias observadas para las señales de agua y metanol y por lo tanto para producir un impulso modulado ajustado para el experimento final.

Parámetros de adquisición

Secuencia de pulso: zgpsd0	Current pulse program
AQ_mod: DQD	Acquisition mode
AUNM: au_prof1d_extract	Acquisition AU program
AUNMP: proc_prof1d_extract pilot.lims	Processing AU program
TD: 65536	Size of fid
NS: 4	Number of scans
DS: 2	Number of dummy scans

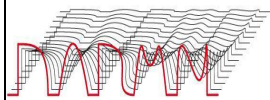
TD0: 1	Loop count for 'td0'
SW [ppm]: 20.5581	Spectral width
SWH [Hz]: 8223.685	Spectral width
AQ [s]: 3.9846	Acquisition time
RG: 4	Receiver gain
DW [us]: 60.800	Dwell time
D1 [s]: 1.0	Delays
D8 [s]: 0.020	Delays
D16 [s]: 0.0002	Delays

Experimento 3: Finalmente, un experimento 1D (noesygpps) se registra, RD – G1 – P(90°) – 4 μ s – P(90°) – mixing time – G2 – P(90°) – AQ, con una ganancia del receptor de 90.5, un tiempo de mezcla de 10 ms, 4 *dummy scans* y 64 FID (aprox. 10 minutos). Todos los demás parámetros se mantienen similares al experimento 2. Los dos gradientes de desenfoque, G1 y G2, de 1 ms cada uno, se añaden para mejorar la supresión de banda selectiva. FID se multiplica por una función exponencial 0.3 Hz (apodización) antes de la transformación de Fourier y se permite la corrección de fase de orden cero.

Parámetros de adquisición

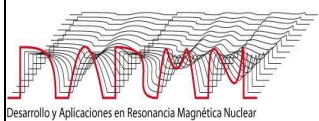
Secuencia de pulso: noesygpps1d.comp	Current pulse program
AQ_mod: DQD	Acquisition mode
AUNM:	Acquisition AU program
AUNMP: proc_prof1d_extract pilot.lims	Processing AU program
TD: 65536	Size of fid
NS: 64	Number of scans
DS: 4	Number of dummy scans
TD0: 1	Loop count for 'td0'
SW [ppm]: 20.5581	Spectral width
SWH [Hz]: 8223.685	Spectral width
AQ [s]: 3.9846	Acquisition time
RG: 90.5	Receiver gain
DW [us]: 60.800	Dwell time
D1 [s]: 1.0	Delays
D8 [s]: 0.020	Delays
D16 [s]: 0.0002	Delays

ANEXO 6.7

 Desarrollo y Aplicaciones en Resonancia Magnética Nuclear	PROTOCOLO PARA EL ANÁLISIS DE EXTRACTOS DE CAFÉ TOSTADO POR GC-MS	Código: COFFEE-v.01 Fecha de actualización: 26/05/2013
--	---	--

Parámetros	Descripción
Columna GC	DB-17MS (30 m, 0.25 mm, 0.25 mm, J&W Scientific, Folsom, CA, USA)
Gas de arrastre	Helio
Flujo del gas de arrastre	1.0 mL/min
Presión columna	7.6 psi
Temperatura de inyección	280 °C
Temperatura de detector	250 °C
Potencial de ionización	70 eV
Rango de masas	50 – 550 amu
Programación temperatura	50 °C, 2 min a 50 °C, 10 °C/min a 220 °C, 5 °C/min a 300 °C, 5 min a 300 °C (Tiempo total de corrida: 40 min)
Volumen de inyección	1 µL
Modo de inyección	Modo split (10:1)

ANEXO 6.8

	PROTOCOLO PARA EL ANÁLISIS DE EXTRACTOS DE CAFÉ TOSTADO POR GC-C-IRMS	Código: COFFEE-v.01 Fecha de actualización: 26/05/2013
---	--	--

Parámetros	Descripción
Columna GC	DB-17MS (30 m, 0.25 mm, 0.25 mm, J&W Scientific, Folsom, CA, USA)
Gas de arrastre	Helio
Flujo del gas de arrastre	1.2 mL/min
Fuente de ionización	Impacto electrónico, 3.0 kV
Temperatura de inyección	280 °C
Temperatura de detector	230 °C
Temperatura del horno de combustión	940 °C
Temperatura del horno de reducción	600 °C
Programación temperatura	70 °C, 2 min a 70 °C, 15 °C/min a 160 °C, 10 °C/min a 280 °C, 2 min a 280 °C (Tiempo total de corrida: 32 min)
Volumen de inyección	1 µL
Modo de inyección	Modo splitless (1.50 min)

ANEXO 6.9

PREDICCIÓN DE LOS MODELOS DE CLASIFICACIÓN PLS-DA

Tabla A6.9a. Predicción del modelo PLS-DA del origen botánico (Or Bot) y geográfico (Or Geo) de los 14 datos espectrales de café verde que conforman el conjunto de validación externa. Se reporta el valor real y el predicho (Pred), así como la frecuencia (Frec) que indica la cantidad de veces que fueron correctamente clasificados cuando se realizaron las predicciones con los 100 modelos

ID	Or Bot Real	Or Bot Pred	Frec	Or Geo Real	Or Geo Pred	Frec
AC1236 CNAOR12067	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC 1240 NBEAH12030	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC 1248 NSJAH12038	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1356 CPIAH12175	Arábica	Arábica	1.00	Colombia	Colombia	0.99
AC1324 UGAAH12153	Arábica	Arábica	1.00	Otro	Otro	0.97
AC1297 GUAH11109	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1307 DOMAH11119	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1289 SALAH11099	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1332 INDRS10071	No Arábica	No Arábica	1.00			
AC1252 TOGORS11070	No Arábica	No Arábica	1.00			
AC1256 IDRH11073	No Arábica	No Arábica	1.00			
AC1293 MEXRH11104	No Arábica	No Arábica	1.00			
AC1259 UGARH11076	No Arábica	No Arábica	1.00			
AC1342 CMFRS10072	No Arábica	No Arábica	1.00			

obtenidos.

Tabla A6.9b. Predicción del modelo PLS-DA del origen botánico (Or Bot) y geográfico (Or Geo) de los 50 datos espectrales de café tostado que conforman el conjunto de validación externa. Se reporta el valor real y el predicho (Pred), así como la frecuencia (Frec) que indica la cantidad de veces que fueron correctamente clasificados cuando se realizaron las predicciones con los 100 modelos obtenidos.

ID	Or Bot Real	Or Bot Pred	Frec	Or Geo Real	Or Geo Pred	Frec
AC1595 CASAH13325	Arábica	Arábica	1.00	Colombia	Colombia	0.93
AC1356 CPIAH12175	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1354 CPIAH12173	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1439 TOLAH13248	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1655 NARAH14391	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1648 SPIAH14381	Arábica	Arábica	1.00	Colombia	Colombia	0.99
AC1245 NSAAH12035	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1199 HACA12052	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1244 NCOAH12034	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1322 SRIAH12151	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1315 SZAAH12144	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1652 NBEAH14353	Arábica	Arábica	1.00	Colombia	Colombia	1.00

AC1206_HPAAH12049	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1589_HPIAH13319	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1323_SBUAH12152	Arábica	Arábica	1.00	Colombia	Colombia	0.84
AC1649_NBEAH14350	Arábica	Arábica	1.00	Colombia	Otro	0.81
AC1635_CENAH_13365	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1591_HPIAH13321	Arábica	Arábica	1.00	Colombia	Colombia	1.00
AC1208_HCOAH12050	Arábica	Arábica	1.00	Colombia	Colombia	0.99
AC1640_CTIAH13373	Arábica	Arábica	1.00	Colombia	Otro	0.84
AC1425_HONAH13237	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1415_BRAA12224	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1328_GUAAH12157	Arábica	Arábica	1.00	Otro	Otro	0.98
AC1365_HAWAS12179	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1501_CRIANN13308	Arábica	Arábica	1.00	Otro	Otro	0.96
AC1412_PERAH12221	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1611_BRAAN13341	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1613_BRAAN13343	Arábica	Arábica	1.00	Otro	Otro	0.99
AC1379_CRIAH12191	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1392_ETIA12204	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1394_GUALH12206	Arábica	Arábica	1.00	Otro	Colombia	0.67
AC1605_TANAH13335	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1687_PANAM14368	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1302_PERAH12114	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1685_PANAH14366	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1674_BRAAN14382	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1327_PERAH12156	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1198_BRAAS12062	Arábica	Arábica	1.00	Otro	Otro	0.98
AC1336_BRAAS12065	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1601_NICAH13331	Arábica	Arábica	1.00	Otro	Otro	1.00
AC1296_UGARS11108	No Arábica	No Arábica	1.00			
UV1204_M3_50	No Arábica	No Arábica	1.00			
AC1342_CMFRS10072	No Arábica	No Arábica	1.00			
UV1257_M11_10	No Arábica	No Arábica	0.85			
UV1210_M4_30	No Arábica	No Arábica	1.00			
AC1417_GUARH12227	No Arábica	No Arábica	1.00			
UV1237_M8_20	No Arábica	No Arábica	1.00			
AC1342_CMFRS10072	No Arábica	No Arábica	1.00			
UV1249_M10_5	No Arábica	No Arábica	0.78			
AC1162_Rove100-16	No Arábica	No Arábica	1.00			

Tabla A6.9c. Predicción del modelo PLS-DA del origen botánico (Or Bot) y geográfico (Or Geo) de 52 datos espectrales correspondientes a 26 muestras de café tostado catalogadas como sospechosas. Se reporta el valor predicho (Pred), así como la frecuencia (Frec) que indica la cantidad de veces que fueron correctamente clasificados cuando se realizaron las predicciones con los 100 modelos obtenidos. Además, se reporta la predicción PLSr del contenido de % Robusta para las muestras clasificadas como No Arábica

ID Muestra	Predicción Or Bot	Frec	Predicción Or Geo	Frec	Pred. PLSr (% Robusta)	Promedio Pred. PLSr (% Robusta)
AC1366_2140	No arábica	1.00			100.77	100.57
	No arábica	1.00			100.35	
	No arábica	1.00			100.58	

AC1367_2285	No árábica	1.00			95.65	95.40
	No árábica	1.00			95.17	
	No árábica	1.00			95.39	
AC1372_2418	Arábica	1.00	Otro	1.00		
	Arábica	1.00	Otro	1.00		
	Arábica	1.00	Otro	1.00		
AC1398_2151uv	No árábica	0.84			60.46	60.95
	No árábica	0.94			61.16	
	No árábica	0.97			61.24	
AC1399_2357uv	No árábica	1.00			44.83	44.51
	No árábica	1.00			44.19	
	No árábica	1.00			44.52	
AC1452_NESTRI271	Arábica	0.97	Otro	0.67		
	Arábica	1.00	Otro	0.67		
	Arábica	1.00	Otro	0.78		
AC1472_13FE002654	No árábica	1.00			81.30	82.03
	No árábica	1.00			82.82	
	No árábica	1.00			81.96	
AC1473_13FB002635	Arábica	1.00	Colombia	0.67		
	Arábica	1.00	Colombia	0.78		
	Arábica	1.00	Colombia	0.75		
AC1474_ZNESNA300	Arábica	0.92	Otro	0.56		
	Arábica	0.91	Otro	0.56		
	Arábica	0.98	Otro	0.64		
AC1475_ZNESTR301	No árábica	1.00			48.05	48.30
	No árábica	1.00			48.57	
	No árábica	1.00			48.28	
AC1476_ZJAMBMO302	Arábica	1.00	Otro	0.89		
	Arábica	1.00	Otro	0.97		
	Arábica	1.00	Otro	0.97		
AC1477_ZPERVRI303	Arábica	1.00	Otro	1.00		
	Arábica	1.00	Otro	0.99		
	Arábica	1.00	Otro	1.00		
AC1478_ZPERCUS304	Arábica	1.00	Otro	0.97		
	Arábica	1.00	Otro	0.98		
	Arábica	1.00	Otro	0.98		
AC1479_ZINDOMA305	Arábica	1.00	Otro	0.70		
	Arábica	1.00	Otro	1.00		
	Arábica	1.00	Otro	0.98		
AC1504_CUNCR13311	Arábica	1.00	Colombia	0.85		
	Arábica	1.00	Colombia	0.78		
	Arábica	1.00	Colombia	0.89		
AC1505_CUNDCR13312	Arábica	1.00	Colombia	1.0		
	Arábica	1.00	Colombia	1.0		
	Arábica	1.00	Colombia	1.0		
AC1506_CUNDCR13313	Arábica	1.00	Colombia	0.85		
	Arábica	1.00	Colombia	1.00		
	Arábica	1.00	Colombia	0.98		
AC1507_CUNDCR13314	Arábica	1.00	Colombia	1.00		
	Arábica	1.00	Colombia	1.00		
	Arábica	1.00	Colombia	1.00		

AC1508_CUNDCR13315	Arábica	1.00	Colombia	1.00		
	Arábica	1.00	Colombia	1.00		
	Arábica	1.00	Colombia	1.00		
AC1581_2782	Arábica	1.00	Otro	1.00		
	Arábica	1.00	Otro	0.89		
	Arábica	1.00	Otro	0.98		
AC1681_2977	Arábica	1.00	Colombia	0.97		
	Arábica	1.00	Colombia	0.98		
	Arábica	1.00	Colombia	0.98		
AC1682_2981	No arábica	1.00			100.43	100.68
	No arábica	1.00			100.87	
	No arábica	1.00			100.74	
AC1683_2985	No arábica	0.80			39.89	40.09
	No arábica	0.85			40.23	
	No arábica	0.92			40.16	
AC1712_14FALM0001	Arábica	0.98	Otro	0.91		
	Arábica	0.97	Otro	0.91		
	Arábica	0.98	Otro	0.98		
AC1713_3097	Arábica	0.99	Otro	0.73		
	Arábica	0.97	Otro	0.67		
	Arábica	0.98	Otro	0.78		
AC1714_3158	Arábica	0.68	Otro	0.98		
	Arábica	0.70	Otro	0.96		
	Arábica	0.79	Otro	0.99		

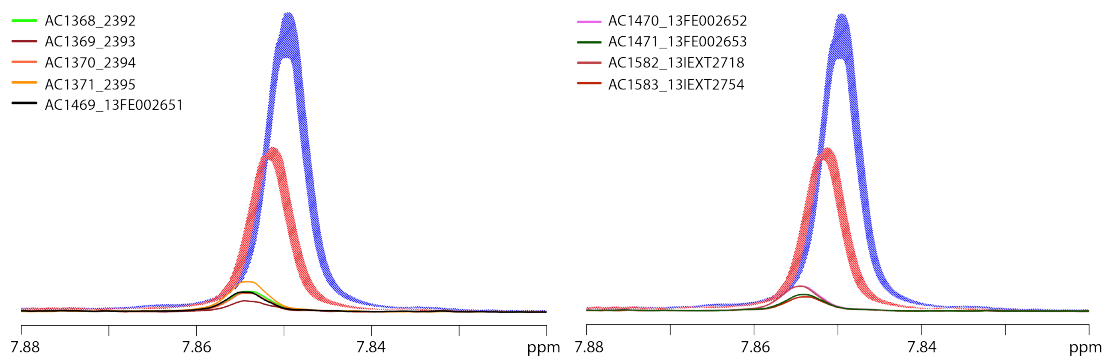


Figura A6.9. Región de la cafeína en el espectro ^1H -RMN del extracto metanólico de las 9 muestras de café tostado catalogadas como sospechosas, en donde se evidencia la ausencia de la señal de la cafeína en estas muestras, comparada con la superposición de las señales de las especies *C. arabica* (rojo) y *C. canephora* (azul).

ANEXO 6.10

Tabla A6.10a. Datos de la relación isotópica de cafeína por GC-C-IRMS de las 34 muestras de café tostado para los tres países de estudio.

País	$\delta^{13}\text{C}$ Cafeína Repetición 1	$\delta^{13}\text{C}$ Cafeína Repetición 2	$\delta^{13}\text{C}$ Cafeína Promedio
Brasil	-28.06	-28.27	-28.17
	-27.95	-27.98	-27.96
	-29.32	-29.50	-29.41
	-29.09	-28.93	-29.01
	-28.93	-28.80	-28.87
	-28.15	-27.91	-28.03
	-30.88	-30.91	-30.89
	-29.82	-29.88	-29.85
	-28.41	-28.44	-28.43
	-29.59	-29.81	-29.70
	-30.27	-29.95	-30.11
Colombia	-28.29	-27.94	-28.11
	-28.93	-28.80	-28.87
	-28.15	-28.02	-28.09
	-28.20	-28.37	-28.29
	-28.41	-28.51	-28.46
	-27.72	-27.74	-27.73
	-28.61	-28.72	-28.66
	-28.50	-28.16	-28.33
	-28.84	-29.21	-29.02
	-28.39	-28.38	-28.38
	-28.23	-28.06	-28.14
	-28.73	-28.89	-28.81
	-29.03	-28.77	-28.90
	-28.58	-28.66	-28.62
	-29.18	-29.11	-29.14
Perú	-29.30	-29.15	-29.22
	-30.03	-30.04	-30.03
	-29.35	-29.27	-29.31
	-30.26	-30.09	-30.18
	-29.64	-30.08	-29.86
	-30.52	-30.25	-30.38
	-30.20	-30.26	-30.23
	-28.53	-28.60	-28.57

Tabla A6.10b. Promedio de los datos de la composición porcentual de los siete compuesto medidos por GC-MS de las 34 muestras de café tostado para los tres países de estudio.

País	2-Furanmetanol	Ácido palmitico	Cafeína	α -tocoferol	β -tocoferol	Estigmasterol	Sitosterol
Brasil	0.69	4.55	74.59	3.45	1.10	0.52	0.47
	0.94	6.42	70.47	4.14	1.21	0.59	0.73
	1.00	3.89	74.40	4.64	1.45	0.60	0.83
	1.26	3.83	72.55	4.73	1.25	0.65	0.70
	1.21	6.50	72.97	4.14	1.23	0.55	0.49
	1.38	4.56	73.83	4.19	1.29	0.63	0.51
	1.06	8.30	75.96	3.60	1.18	0.74	0.95
	1.28	6.27	70.86	4.44	1.31	0.69	0.66
	1.06	4.26	74.97	3.94	1.28	0.55	0.65
	1.12	10.17	75.87	2.90	1.10	0.52	0.74
	1.03	4.64	74.61	3.68	1.14	0.48	0.49
Colombia	1.12	5.99	70.01	4.28	1.14	0.53	0.52
	1.14	8.57	68.87	3.02	1.07	0.69	0.76
	0.98	6.34	70.91	3.60	0.98	0.38	0.51
	0.96	6.25	71.58	3.35	1.11	0.55	0.86
	1.20	4.37	70.23	3.73	1.16	0.63	0.54
	0.89	7.73	70.06	3.06	0.97	0.64	0.63
	1.12	7.30	68.92	3.19	1.04	0.67	0.82
	1.18	6.62	66.77	4.52	1.20	0.59	0.76
	1.13	6.63	68.35	3.84	1.08	0.47	0.51
	1.18	6.96	68.42	3.58	1.04	0.57	0.49
	1.32	5.96	68.01	4.22	1.23	0.61	0.86
	1.16	6.20	69.00	3.18	0.99	0.66	0.76
	1.26	6.64	69.89	3.28	0.84	0.53	0.59
	1.09	6.50	67.43	4.10	1.18	0.63	0.55
	1.17	4.26	70.19	4.14	1.22	0.72	0.83
Perú	1.17	3.19	72.30	4.34	1.23	0.46	0.59
	1.39	4.18	72.84	2.84	1.00	0.63	0.56
	1.47	4.77	73.37	4.08	1.23	0.50	0.49
	1.24	9.05	70.87	3.15	1.11	0.50	0.62
	1.30	5.16	72.10	4.61	1.33	0.51	0.61
	1.27	4.07	73.79	3.70	1.12	0.48	0.38
	1.13	8.78	75.44	3.45	1.25	0.78	0.92
	1.32	7.15	72.35	3.76	1.08	0.45	0.46

ANEXO 6.11

PUBLICACIONES Y COMUNICACIONES A CONGRESOS

PUBLICACIONES

Arana, V.A., Medina, J., Alarcon, R., Moreno, E., Heintz, L., Schäfer, H., Wist, J. (2015). Coffee's country of origin determined by NMR: The Colombian case. *Food Chemistry*, 175, 500-506. ISSN 0308-8146.

Arana, V.A., Medina, J., Alarcón, R., Moreno, E., Wist, J. (2012). ¹H-NMR Profiling of Green and Roasted Coffee Extracts: Classification of Colombian Coffees. *24th International Conference on Coffee Science*. p. 172-178. ISBN 978-2-900212-23-3.

COMUNICACIONES A CONGRESOS NACIONALES E INTERNACIONALES

Julien Wist, Victoria Andrea Arana, Jessica Medina. Coffee fingerprinting by NMR. **4th Annual Practical Applications of NMR in Industry Conference**. San Diego, Estados Unidos de América. Febrero 9-12, 2015. *Comunicación oral*.

Victoria Andrea Arana, Jessica Medina, Rodrigo Alarcón, Edgar Moreno, Lea Heintz, Hartmut Schaefer, Julien Wist. Determination of regional geographic origin of roasted coffee using GC-C-IRMS, GC-MS and ¹H-NMR. **25th International Conference on Coffee Science**. Armenia, Colombia. Septiembre 8-13, 2014. *Poster*.

Victoria Andrea Arana, Jessica Medina, Rodrigo Alarcón, Edgar Moreno, Lea Heintz, Hartmut Schaefer, Julien Wist. ¹H-NMR fingerprinting and GC-MS profiling of roasted coffee for geographical origin. **XII Congreso Colombiano de Fitoquímica**. Cartagena, Colombia. Mayo 21-23, 2014. *Poster*.

Jessica Medina, Victoria Andrea Arana, Rodrigo Alarcón, Edgar Moreno, Julien Wist. Las moléculas que hacen el café colombiano. **III Simposio de Química**. Cali, Colombia. Octubre 29-Noviembre 1, 2013. *Poster*.

Jessica Medina, Victoria Andrea Arana, Rodrigo Alarcón, Edgar Moreno, Julien Wist. ^1H -RMN fingerprinting para la detección de adulteraciones en café. **VIII Simposio Colombiano de Catálisis - VIII SICCAT y VI Simposio de Química Aplicada - SIQUIA 2013**. Armenia, Colombia. Septiembre 11-13, 2013. *Poster*.

Victoria Andrea Arana, Jessica Medina, Rodrigo Alarcón, Edgar Moreno, Lea Heintz, Hartmut Schaefer, Julien Wist. Detection of Adulterated Coffee by ^1H -NMR and Chemometrics Methods. **9th EUROMAR 2013 conference: a European Magnetic Resonance Meeting**. Creta, Grecia. Junio 30-Julio 5, 2013. *Poster*.

Victoria Andrea Arana, Jessica Medina, Rodrigo Alarcón, Edgar Moreno, Julien Wist. ^1H -NMR Profiling of Green and Roasted Coffee Extracts: Classification of Colombian Coffees. **24th International Conference on Coffee Science**. San José, Costa Rica. Septiembre 8-12, 2012. *Comunicación oral*.

Victoria Andrea Arana, Jessica Medina, Rodrigo Alarcón, Edgar Moreno, Lalita Uribe, Andrés Mauricio Castillo, Luc Patiny, Julien Wist. Fully automated classification of Colombian Coffee using NMR. **IV Ibero- American NMR Meeting, IV GERMN Bienal Meeting, III Iberian NMR Meeting**. Aveiro, Portugal. Septiembre 25-28, 2012. *Comunicación oral*.