



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

“CARACTERIZACIÓN
METABOLÓMICA DE 5 ACCESIONES
DE AJÍES NATIVOS (*CAPSICUM* SPP.)
DE UNA COLECCIÓN PERUANA COMO
HERRAMIENTA PARA LA
DISCRIMINACIÓN DE LAS
ACCESIONES Y LA PREDICCIÓN DE
SUS ATRIBUTOS SENSORIALES”.

TESIS PARA OPTAR EL GRADO DE
DOCTOR EN CIENCIAS DE LA VIDA

FABIO FRANCISCO ESPICHAN

JAUREGUI

LIMA-PERÚ

2020

ASESOR

PhD., ROSARIO ROJAS DURÁN

CO ASESOR

PhD., GUILLAUME MARTI

JURADO DE TESIS

PhD. LUIS JULIO CÉSAR DE STEFANO BELTRÁN

PRESIDENTE

PhD. LUZ AURORA CARBAJAL ARROYO

PhD. ANA LUCY SICCHA MACASSI

VOCAL

PhD. FRANCISCO JOSÉ PEIRANO BLONDET

SECRETARIO

DEDICATORIA

Esta tesis está dedicada a Dios mi salvador quien me sostiene. A mi papá Fabio Espichán Aguirre quien me apoyo en todo. Para Aquellos que me apoyaron desinteresadamente y aportaron un granito de arena en este trabajo.

AGRADECIMIENTOS

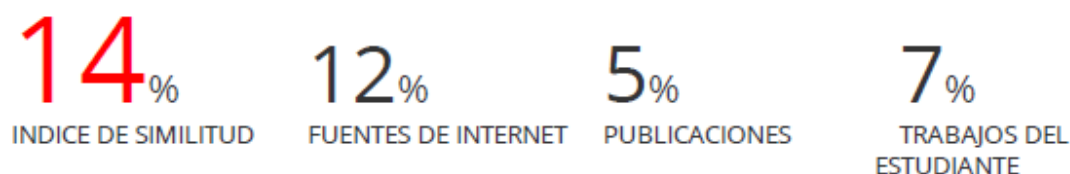
Agradezco a Dios, por ser mi fortaleza. A mis padres y hermanos y amigos que estuvieron conmigo apoyándome en cada momento y durante el desarrollo de este trabajo de tesis. Gracias a todos y que Dios les bendiga.

FUENTES DE FINANCIAMIENTO

becario de la escuela doctoral franco - peruana en ciencias de la vida

CARACTERIZACIÓN METABOLÓMICA DE 5 ACCESIONES DE AJÍES NATIVOS (CAPSICUM SPP.) DE UNA COLECCIÓN PERUANA COMO HERRAMIENTA PARA LA DISCRIMINACIÓN DE LAS ACCESIONES Y LA PREDICCIÓN DE SUS ATRIBUTOS SENSO

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	link.springer.com Fuente de Internet	1%
2	webbook.nist.gov Fuente de Internet	<1%
3	hdl.handle.net Fuente de Internet	<1%
4	tel.archives-ouvertes.fr Fuente de Internet	<1%
5	git.embl.de Fuente de Internet	<1%
6	www.mdpi.com Fuente de Internet	<1%
7	journals.plos.org Fuente de Internet	<1%
8	www.frontiersin.org Fuente de Internet	<1%

CONTENIDO

1.	INTRODUCCIÓN	1
2.	PLANTEAMIENTO DEL PROBLEMA	8
3.	MARCO TEÓRICO	11
3.1.	Metabolómica	11
3.2.	Métodos de Análisis multivariado	13
3.2.1.	Análisis de Componentes Principales (PCA).....	13
3.2.2.	Mínimos cuadrados Parciales (PLS)	15
3.2.3.	Análisis discriminante de mínimos cuadrados parciales ortogonales	16
3.2.4.	Validación cruzada.....	18
3.2.5.	R^2 y Q^2	20
3.2.6.	S-plot.....	21
3.3.	Métodos de tratamiento de datos.....	21
3.3.1.	Normalización.....	21
3.3.2.	Escalado	22
3.4.	Análisis factorial múltiple (AFM).....	23
3.4.1.	Objetivos de un AFM.....	24
3.5.	Diagramas Heat Map	25
3.6.	CV-ANOVA	26
3.7.	Curvas ROC (Característica Operativa del Receptor).....	28
3.8.	Combinación de tablas de datos ESI (+, -).....	31
3.9.	Análisis multibloque	34
3.10.	Métodos de análisis Univariado	36
3.10.1.	Análisis de varianza (ANOVA)	36
3.10.2.	Análisis Post-hoc, Test de Tukey	36
3.10.3.	Perfil sensorial	36
3.11.	Bases de datos de Productos Naturales	37

3.11.1.	KNAPSAcK.....	37
3.11.2.	Northern African Natural Products Database (NANPDB).....	37
3.11.3.	Plant Metabolic Network (PMN, PlantCyc)	38
3.11.4.	Universal Natural Products Database (UNPD)	38
3.11.5.	Chemical Entities of Biological Interest (ChEBI).....	39
3.11.6.	LipidMAPS.....	39
3.11.7.	Dictionary of Natural Products (DNP).....	39
3.11.8.	NIST 14.....	40
3.11.9.	Global Natural Products Social Molecular Networking (GNPS).....	40
3.12.	Ajíes Nativos Peruanos (<i>Capsicum</i> spp.)	41
3.13.	Taxonomía del género <i>Capsicum</i>	42
3.14.	Ajíes del Perú, accesiones de la colección del INIA y UNALM	44
3.15.	El gusto, sabor y aroma.....	47
3.16.	Análisis descriptivo sensorial.....	49
4.	JUSTIFICACIÓN DEL ESTUDIO	51
5.	OBJETIVOS	54
5.1.	Objetivo principal	54
5.2.	Objetivos específicos	54
6.	PROCEDIMIENTO EXPERIMENTAL	56
6.1.	Diseño experimental	56
6.2.	Población	57
6.3.	Muestra	58
6.4.	Operacionalización de variables	58
6.5.	Procedimiento y técnicas	59
6.6.	Consideraciones éticas	59
6.7.	Flujo de trabajo de identificación de Biomarcadores.....	62
6.8.	Flujo de trabajo de validación de Biomarcadores	63
6.9.	Flujo de trabajo para el análisis multibloque	64

6.10.	Materiales y Métodos.....	66
6.10.1.	Material de la Planta	66
6.10.2.	Liofilización de las muestras.....	66
6.10.3.	Análisis sensorial	67
6.10.4.	Análisis proximal	71
6.10.5.	Determinación de Humedad.....	71
6.10.6.	Determinación de grasas	72
6.10.8.	Determinaciones Instrumentales	76
6.10.9.	Análisis estadístico.....	76
6.10.10.	Análisis por UHPLC-HRMS Orbitrap.....	77
6.10.11.	Perfilado UHPLC-HRMS Orbitrap	78
6.10.12.	Adquisición de la data	81
6.10.13.	Procesamiento de Data UHPLC-HRMS.....	81
6.10.14.	Análisis por SPME-GC-MS	82
6.10.15.	Adquisición y procesamiento de la data SPME-GC-MS	83
7.	RESULTADOS	85
7.1.	Perfil de metabolitos no volátiles de los ajíes de Chincha	85
7.2.	Perfil de metabolitos volátiles de los ajíes de Chincha	94
7.3.	Tratamiento de la data sensorial.....	102
7.3.1.	Evaluación de la Normalidad de la data.....	102
7.3.2.	Análisis del panel de degustación	108
7.3.3.	Análisis proximal e instrumental	110
7.3.4.	Optimización del modelo del perfil sensorial	111
7.3.5.	Evaluación de la calidad del modelo sensorial optimizado.....	111
7.3.6.	Estimación de la confiabilidad del modelo sensorial	113
7.3.7.	AFM de la data sensorial optimizada.....	122
7.3.8.	Correlación de las variables sensoriales y fisicoquímicas.....	132
7.3.9.	Normalización de la data sensorial optimizada de Chincha.....	137

7.3.10.	AFM de la data sensorial optimizada y normalizada de Chincha	143
7.3.11.	Diagramas Heatmap y PLS-DA de la data sensorial de Chincha.....	147
7.3.12.	Perfil sensorial, diagramas de barras.....	155
7.4.	Análisis multivariado de metabolitos volátiles	161
7.5.	Diferencias Metabólicas entre cosecha Orgánica y Concencional.....	177
7.6.	Discriminación de accesiones y especies y su huella metabólica	186
7.6.1.	Discriminación de especies de los ajés de Chincha	186
7.6.2.	Discriminación de accesiones de los ajés de Chincha.....	212
7.7.	Redes moleculares y clases químicas de metabolitos	233
7.8.	Validación del modelo de huella metabólica	242
7.9.	Identificación de descriptores químicos del sabor y aroma.....	258
7.9.1.	Análisis multibloque	258
8.	DISCUSIÓN DE RESULTADOS.....	285
8.1.	Analisis sensorial de las cinco accesiones de ajés nativos.....	285
8.3.	Metabólica de las cinco accesiones de ajés nativos, biomarcadores	297
8.4.	Análisis multivariado de la data SPME-GC-MS, ajé fresco y liofilizado.....	307
8.5.	Redes Moleculares como herramienta en la identificación de biomarcadores	313
8.6.	Validación del modelo de huella metabólica, biomarcadores	314
8.7.	Análisis Multibloque, correlaciones.....	323
8.8.	Perspectivas Metabólicas con enfoque hacia metabolitos desconocidos	333
9.	CONCLUSIONES	335
10.	RECOMENDACIONES.....	337
12.	BIBLIOGRAFÍA	338
13.	ANEXO	
13.1.	ANEXO 1 (Perfil de metabolitos No volátiles)	
13.2.	ANEXO 2 (Perfil de Metabolitos Volátiles)	

RESUMEN

Se ha caracterizado sensorialmente cinco accesiones de ajíes nativos (Ayuclo, Tomatito rojo, Charapita, Miscucho y Dulce rojo cultivados en las localidades de Huaral, Chiclayo y Chincha) con un panel de cata entrenado utilizando 6 descriptores del aroma y 4 descriptores del sabor. Con la data sensorial, se desarrolló un modelo estadístico sensorial que ha permitido clasificar y discriminar a las accesiones de los ajíes nativos. Por otro lado, se determinó el perfil metabolómico de las cinco accesiones de ajíes nativos utilizando técnicas cromatográficas de alto rendimiento (UHPLC-HRMS Orbitrap y GC-MS). También, en base a los perfiles metabolómicos y mediante herramientas estadísticas de análisis multivariado y uso de software especializados, se ha desarrollado un modelo estadístico de huella metabólica que ha permitido la identificación de los biomarcadores putativos que permiten discriminar las accesiones y las especies de genotipos de *Capsicum*. Utilizando estos resultados se aplicó un novedoso enfoque de análisis multivariado multibloque que ha permitido vincular los metabolitos del sabor (metabolitos no volátiles) y metabolitos del olor (metabolitos volátiles) con los scores del sabor y aroma dados por el panel de cata. De esta manera se han identificado los descriptores químicos o metabolitos que permiten explicar el sabor y aroma de los ajíes nativos.

Palabras clave: Metabolómica, *Capsicum*, Análisis multivariado, Análisis sensorial, ajíes nativos Peruanos.

ABSTRACT

Five accessions of native peppers (Ayucello, Tomatito rojo, Charapita, Miscucho and Dulce rojo cultivated in the towns of Huaral, Chiclayo and Chincha) have been sensory characterized with a trained tasting panel using 6 aroma descriptors and 4 flavor descriptors. With the sensory data, a sensory statistical model was developed that has allowed classifying and discriminating the accessions of native peppers. On the other hand, the metabolomic profile of the five accessions of native peppers was determined using high performance chromatographic techniques (UHPLC-HRMS Orbitrap and GC-MS). Also, based on metabolomic profiles and through statistical tools of multivariate analysis and use of specialized software, a statistical model of metabolic footprint has been developed that has allowed the identification of putative biomarkers that allow discriminating accessions and genotype species of *Capsicum*. Using these results, a novel multivariate multiblock analysis approach was applied that allowed us to link taste metabolites (non-volatile metabolites) and odor metabolites (volatile metabolites) with taste and aroma scores given by the tasting panel. In this way, chemical descriptors or metabolites have been identified that explain the taste and aroma of native peppers.

Keywords: Metabolomics, *Capsicum*, Multivariate analysis, Sensory analysis, native Peruvian peppers.

Índice de abreviaturas

AFM = Análisis factorial múltiple

AMV = análisis multivariado

DA = análisis discriminante

DIABLO= Data Integration Analysis for Biomarker discovery using a Latent component method for Omics studies

GC-MS = cromatografía de gases acoplada a espectrometría de masas

HCA = análisis jerárquico de clúster

INIA = Instituto Nacional de Innovación Agraria

LDA = análisis discriminante lineal

OPLS-DA = Análisis discriminante de proyecciones ortogonales de estructuras latentes

PCA = análisis de componentes principales

PLS = Mínimos cuadrados parciales

SPME= microextracción en fase sólida

QDA = análisis descriptivo cuantitativo

S-PLS-DA= Sparce-PLS-DA

UHPLC-HRMS = cromatografía líquida de ultra alta performance acoplada a espectrometría de masas de alta resolución.

UNALM = Universidad Nacional Agraria La Molina

UPCH = Universidad Peruana Cayetano Heredia

1. INTRODUCCIÓN

El ají (*Capsicum spp.*) de la familia Solanaceae, es uno de los cultivos más importantes del mundo (1). Su cultivo está muy extendido debido a su uso como especia para realzar el sabor de las comidas (2). Los frutos de *Capsicum* son fuente de muchos compuestos beneficiosos para la salud tales como ácido ascórbico (vitaminas C), tocoferoles (Vitamina E), carotenoides (provitamina A), flavonoides (antioxidantes) y capsaicinoides; estos últimos compuestos son responsables de su pungencia (3). El género *Capsicum* ofrece muchas variedades y diversidad de colores, tamaños, pungencia, aromas y sabores en las cocinas de la mayor parte del mundo (4). Algunas son variedades de gran importancia económica a nivel mundial, como el ají paprika (*C. annuum*) (5).

El género *Capsicum* exhibe una variación morfológica considerable, especialmente en las formas de los frutos, colores y tamaños. Muchas especies tienen estados de carácter morfológico superpuestos, llevando a la identificación de especies no resueltas o erróneas. Generalmente se requiere una combinación de caracteres de diagnóstico para identificar y diferenciar las especies del género *Capsicum* (6).

Aunque el género *Capsicum* ha sido estudiado desde el punto de vista taxonómico y evolutivo, utilizando los caracteres morfológicos, citológicos y moleculares, sin embargo, aún existen algunos problemas relacionados a la delimitación taxonómica del género y sus especies, la nomenclatura de las formas silvestres y cultivadas, y

el tratamiento de variación infraespecífica. Los taxa *C. annuum*, *C. chinense* y *C. frutescens* forman un complejo poco diferenciado y no está suficientemente definido si constituyen o no especies distintas (7). A pesar de la confusión que rodea a la clasificación, nomenclatura e identificación de las especies en *Capsicum*, se acepta que hay alrededor de 38 especies identificadas, de las que *Capsicum annuum*, *C. chinense*, *C. frutescens*, *C. baccatum* y *C. pubescens* están domesticadas (8,9). Las especies *C. annuum*, *C. chinense*, *C. baccatum* y *C. frutescens* pueden ser entrecruzadas y producir híbridos fértiles (10).

Para resolver los problemas detrás de la taxonomía se han desarrollado estrategias moleculares que permiten en cierto modo acelerar el desarrollo del conocimiento taxonómico de las especies. En ese sentido, los enfoques de taxonomía molecular se han hecho patentes en la solución de problemas específicos dentro de la taxonomía, ya que se centran en datos derivados de material genético (11,12). Varios estudios demuestran que, en comparación con el criterio morfológico tradicional para delimitar las especies, las herramientas moleculares a veces permiten la detección de especies adicionales, llamadas crípticas (2 o más especies distintas clasificadas como una sola especie.), que no podrían distinguirse únicamente por motivos morfológicos (13,14). En la actualidad, la identificación inequívoca de los cultivares de plantas, la resolución de controversias taxonómicas y seguimiento de la diversidad genética dentro y entre cultivares se realiza a través de marcadores moleculares basados en PCR de un gran número de cultivos importantes (15–17). En el género *Capsicum* se han hecho algunos avances

mediante caracterización molecular vía AFLP, SSR (microsatélites) y Citogenética **(18–20)**.

En la continuación del dogma central de la biología molecular, la cascada ómica inicia con la genómica, seguida de la proteómica y finaliza con la metabolómica. Esta última, por definición es el campo de investigación multidisciplinario concerniente con el estudio de los metabolomas, de los metabolitos presentes naturalmente y exógenos (por ejemplo, contaminantes ambientales), de bajo peso molecular (típicamente <2000 Da) dentro de los sistemas biológicos **(21)**.

La metabolómica puede tener dos enfoques: “targeted” y “non targeted”. El enfoque targeted implica la medición de una lista de determinados metabolitos presentes en muestras crudas, la identidad de dichos metabolitos es conocida, por lo que se pueden estudiar de manera cualitativa y cuantitativa. Dicha lista de metabolitos puede provenir, por ejemplo, de una base de datos de desórdenes de una enfermedad en donde se monitorea la eficacia de una estrategia terapéutica particular y donde el enfoque está en ciertos metabolitos **(22)**. Los enfoques metabolómicos targeted son comúnmente conducidos por una cuestión bioquímica específica o hipótesis que motiva la investigación. En este enfoque se busca comprobar una hipótesis **(23,24)**, y se tiene conocimiento del metabolito en estudio. Tiene una alta fiabilidad cuantitativa, baja complejidad del tratamiento de la data, requiere un alto esfuerzo para la construcción del método y tiene una baja cobertura de metabolitos **(25)**.

En cambio, en la metabolómica non-targeted se realiza un screening de todos los metabolitos conocidos y desconocidos en un determinado extracto crudo, este screening también es llamado perfilado exhaustivo. En metabolómica non-targeted el uso de estadística multivariada, como PCA (análisis de componentes principales), es algo rutinario para presentar los resultados. Este enfoque nos permite generar hipótesis al observar diferencias o similitudes entre las muestras mediante el análisis multivariado (26,27). En ese sentido, la metabolómica non-targeted es la herramienta más adecuada para el análisis exploratorio en la búsqueda de una huella metabólica que permita la discriminación de accesiones y especies en el género *Capsicum*. Entre las aplicaciones exitosas de la metabolómica non-targeted, se puede mencionar el trabajo de Li Jing *et al.*, donde se estudiaron los aceites esenciales de 20 accesiones del género *Citrus*. Los resultados demostraron claramente que los perfiles de metabolitos de *Citrus* podrían utilizarse para la clasificación taxonómica del género y son complementarios a la evidencia taxonómica existente, especialmente para la identificación y diferenciación de especies híbridas (28). Para el género *Capsicum* se ha reportado el trabajo de Yuni Wahyuni *et al.* en el que se realizó un análisis metabolómico y análisis de marcadores moleculares para explorar la diversidad de ajíes en el mundo (29). Se seleccionaron 32 especies de acuerdo a su variación en caracteres morfológicos, pungencia y origen geográfico de muestras de *C. annuum* (16 variedades domesticadas y una especie silvestre de Jamaica), de *C. chinense* (7 variedades domesticadas y 2 variedades peruanas de denominación no determinada), de *C. frutescens* (2 variedades de Indonesia y USA domesticadas) y de *C. baccatum* (3

variedades domesticadas y una variedad silvestre de Bolivia). A partir de esta primera identificación, el análisis genómico con marcadores moleculares de AFLP (polimorfismo de longitud de fragmento amplificado) confirmó el agrupamiento filogenético de las accesiones según las especies, logrando delimitar las 4 especies por AFLP con análisis jerárquico de Cluster (HCA). En contraste con el perfilado metabolómico de los metabolitos semi-polares, estos últimos también pudieron hacer la diferenciación del grupo de especies estudiadas. También se observó que una accesión que inicialmente fue clasificada taxonómicamente como *C. chinense*, mediante AFLP y análisis de Heatmap (mapa jerárquico de cluster) de la data metabolómica fue clasificada finalmente como *C. annuum*. Se observó también del Heatmap que la metabolómica puede discriminar especies en función de su perfil de metabolitos.

En este trabajo se concluyó que la combinación de la genética y la información fenotípica, tal como el perfil de metabolitos, puede ayudar a los mejoradores a desarrollar nuevas variedades con composiciones de metabolitos específicos, deseados y con rasgos relacionados y que serviría como una estrategia de orientación dentro de los programas de mejoramiento de cultivos de la diversidad fenotípica, para la obtención de rasgos de calidad de ajíes importantes (30).

Actualmente, el Instituto Nacional de Innovación Agraria (INIA) y la Universidad Nacional Agraria La Molina (UNALM) han consolidado una colección nacional de Germoplasma de Ajíes oriundos del Perú, en el marco del proyecto “Consolidación

de las colecciones INIA y UNALM de *Capsicum* nativos del Perú con una base de datos morfológica, genética, físico-química y culinaria integrada y análisis de las prioridades de conservación y desarrollo". Las colecciones de ajíes del INIA y la UNALM en conjunto conservan más de 900 accesiones (muestras genéticas). El INIA cuenta con 413 accesiones de ajíes y 296 de rocotos; mientras que la UNALM tiene más de 200 accesiones de ají, colectadas en distintas regiones del Perú, donde se harán estudios morfológicos, genéticos y fisicoquímicos, mas no metabolómicos (31,32). En ese sentido, los estudios metabolómicos contribuirán a la valorización del germoplasma. Del perfil metabolómico de las accesiones estudiadas se podrá caracterizar y poner en valor a través de la composición de metabolitos para su inclusión en los programas de mejoramiento genético y desarrollo de nuevos cultivares. El enfoque metabolómico también podría servir para una caracterización mayor de accesiones y también como herramienta alternativa o complementaria a las herramientas moleculares para la discriminación de accesiones y especies de ajíes nativos.

Por otro lado, además de la caracterización de ajíes en base a criterios morfológicos, genéticos y metabolómicos, también es importante caracterizar los ajíes nativos en base a sus atributos sensoriales (sabor y aroma). Para realizar la caracterización sensorial existen diferentes enfoques, entre ellos está el método de análisis descriptivo cuantitativo o QDA. El QDA se basa en el principio de la capacidad de un panelista para verbalizar las percepciones de un producto de una manera confiable y reproducible y dar una descripción cuantitativa e integral del producto

susceptible a un análisis estadístico (33). Entre los ejemplos que se pueden citar en análisis descriptivo sensorial de alimentos se tiene la caracterización de cacao, vino, queso y pimiento, entre otros (34–37).

Del perfil sensorial de un alimento, es importante destacar también qué metabolitos son los responsables del sabor y aroma. La metabolómica junto al análisis sensorial y análisis multivariado permitirá la caracterización integral de los ajíes nativos y tener un método alternativo o complementario al QDA que pueda servir para una selección objetiva de aromas y sabores cuyas accesiones de la cual provengan puedan ser incluidas en los programas de mejoramiento del género *Capsicum*.

El presente trabajo contribuirá con la caracterización metabolómica y la identificación de los descriptores químicos del sabor y aroma de 5 accesiones de ajíes (Ayuello (*Capsicum baccatum*), Miscucho (*C. chinense*), Tomatito rojo (*C. baccatum*), Charapita amarillo (*C. frutescens*) y Dulce rojo (*C. chinense*)) pertenecientes a la colección de ajíes nativos del INIA.

2. PLANTEAMIENTO DEL PROBLEMA

El sistema de clasificación tradicional taxonómico se basa en gran medida en las propiedades fenotípicas. Los taxones se describen en base a su similaridad morfológica, fisiológica y propiedades bioquímicas (11). Sin embargo, la discriminación de especies por este sistema en algunos casos es complicada debido a que se basa únicamente en caracteres morfológicos, cuya expresión puede verse influenciada por factores ambientales y de desarrollo debido a la interacción del genoma con el medio ambiente y la adaptación (la plasticidad del genoma) (38), dando lugar a identificaciones erróneas. Este es el caso del género *Capsicum*. Las especies *C. annuum*, *C. chinense* y *C. frutescens* están incluidas dentro del grupo de las flores blancas y conforman el complejo taxonómico de *C. annuum*, donde existen barreras de cruce poco desarrolladas y grandes similitudes con respecto a la morfología, isoenzimas y las secuencias de ADN, lo que sugiere que podrían ser de la misma especie (19).

Las herramientas moleculares han logrado efectivamente discriminar entre las especies del género *Capsicum*, sin embargo, no tienen el alcance que ofrece la metabolómica (39,40). El perfil de metabolitos secundarios, un descriptor taxonómico alternativo que no se utiliza con mucha frecuencia, está íntimamente ligado a las características fenotípicas de los ajíes nativos (41). La metabolómica puede realizar una caracterización de especies por medio de la información del perfil

de metabolitos y también diferenciarlas al identificar los biomarcadores, utilizando herramientas de estadística multivariada (42).

El INIA cuenta con 413 accesiones de ajíes y 296 de rocotos; mientras que la UNALM tiene 369 accesiones de ají (31,32,43). Hasta el momento, ambas instituciones tienen problemas en relación con la discriminación de especies de ajíes silvestres de sus colecciones.

En el presente proyecto se estudiará la diferenciación de especies del género *Capsicum* de 5 accesiones de ajíes nativos de la colección del INIA, cuyos nombres comunes son: Ayucello (*Capsicum baccatum*), Miscucho (*C. chinense*), Tomatito rojo (*C. baccatum*), Charapita amarillo (*C. frutescens*) y Dulce rojo (*C. chinense*) desde un enfoque metabolómico. El aporte de este estudio será la utilización de la metabolómica como una herramienta complementaria a la información genética disponible, con miras de su aplicación para una caracterización integral de una colección más grande de accesiones de ajíes nativos del INIA.

Otras características fenotípicas importantes de los ajíes son el sabor y aroma. Las 5 accesiones que se van a evaluar en este proyecto ya han sido analizadas sensorialmente por un panel de cata entrenado en el marco del proyecto “Ajíes Nativos Peruanos, caracterización Agro-morfológica, Químico-Nutricional y sensorial” (44), y fueron seleccionadas por sus atributos notables.

Para conocer el perfil sensorial de los ajíes se utiliza un enfoque de análisis descriptivo cuantitativo o QDA que es una técnica analítica para la caracterización sensorial y control de calidad de los ajíes nativos (44). Sin embargo, este método puede estar sujeto a errores debido a las propias subjetividades de los panelistas. Además, preparar un panel capacitado requiere de un proceso de selección de personal, tiempo y un presupuesto dedicado para tal fin. Mantener el panel activo es importante para obtener resultados reproducibles, de otro modo si no se mantiene se podría “descalibrar”. Por tal motivo se deben buscar métodos alternativos o complementarios al análisis por paneles, para no depender totalmente de esta técnica analítica tan difícil de mantener. Para ello, sería de mucha utilidad saber qué metabolitos se asocian con los atributos sensoriales del sabor y aroma dados por el panel de cata y así poder establecer un método de caracterización objetiva de sabores y aromas en ajíes. La asociación de los metabolitos volátiles (olor) y no volátiles (gusto) con la data sensorial no es algo trivial. Para tal fin, se debe buscar una estrategia o modelo estadístico que se ajuste mejor a las características multivariadas de la data ómica con la data sensorial. El presente estudio se enfocará en la identificación de los metabolitos que expliquen el sabor y aroma de 5 accesiones de ajíes nativos de la colección del INIA a través del enfoque metabolómico. El modelo podrá ser útil para realizar la identificación de metabolitos del sabor y aroma de un mayor número de accesiones de ajíes nativos del INIA.

3. MARCO TEÓRICO

3.1. Metabolómica

La metabolómica es el campo de investigación multidisciplinario de los metabolomas. Los metabolomas están constituidos por metabolitos precursores, intermediarios y productos de rutas bioquímicas, siendo los metabolitos secundarios algunos de los productos terminales de los procesos celulares superiores que provee colectivamente una "huella digital" de la compleja interacción entre el genoma con el medioambiente. Desde una perspectiva analítica, tanto la medición como la identificación de metabolomas completos representa un desafío considerable, debido a la vasta heterogeneidad estructural de los metabolitos, por su gran número (por ejemplo, un estimado de 200,000 metabolitos secundarios estructuralmente distintos en todo el reino vegetal) y sus amplios rangos de concentración (21). La metabolómica involucra el perfilado cualitativo completo de metabolitos en un organismo, utilizando tecnologías analíticas de última generación. Permite examinar la variación total en los perfiles de los metabolitos y puede detectar cambios biológicos complejos utilizando estadística multivariada.

La metabolómica involucra un enfoque eficiente de screening (45). Los enfoques de perfilado no targeted de alto rendimiento se han utilizado para obtener una visión general de la diversidad de metabolitos en colecciones de germoplasma, tales

como Arabidopsis (46), tomate (47) y ají (48). Debido a que estos métodos de alto rendimiento generan un extenso conjunto de datos, se requieren unas técnicas de análisis y visualización especiales para extraer la información relevante y poder elucidar por ejemplo la función de las proteínas y de los metabolitos, las interacciones entre estas moléculas y los mecanismos reguladores subyacentes, etc. El análisis estadístico de tales datos no es algo trivial ya que en muchos casos el número de metabolitos supera al número de muestras. Esto ha dado lugar a un nuevo estándar en el diseño de experimentos en metabolómica: el así llamado paradigma $p > n$. Bajo este nuevo paradigma, el número de variables independientes n o individuos (por ejemplo, muestras de frutas) es mucho más pequeño que el número de variables p (por ejemplo, el número de metabolitos en perfiles GC-MS) que es analizado. Mientras que en configuraciones clásicas se evalúan pocas hipótesis nulas preespecificas, ahora es necesario realizar cientos o miles de pruebas de hipótesis. Los métodos clásicos estadísticos típicamente requieren que el número de variables independientes sea grande y un múltiplo del número de variables, en primer lugar para evitar la colinealidad y sobreajuste (49). Sin embargo, el número de muestras que pueden ser consideradas para experimentos metabolómicos de alto rendimiento es a menudo limitado debido a las técnicas y limitaciones económicas del experimento. Aunque el conocimiento sobre la regulación de la formación de metabolitos está en crecimiento, para miles de metabolitos, su función en la planta, sus pasos biosintéticos y su misma regulación es aún desconocida. La metabolómica en combinación con análisis estadístico ayuda a comprender las vías metabólicas de los principales contrastes metabólicos entre genotipos y en

combinación con análisis genético, puede identificar los QTL (quantitative trait locus) y los genes de los pasos fundamentales clave en las vías metabólicas. Además, el perfilado metabólico permite la detección y subsecuente identificación de compuestos desconocidos correlacionados con un rasgo de interés. Aunque el costo y el grado de heredabilidad de los metabolitos estudiados deben ser tomados en cuenta, la metabolómica como herramienta aplicada a la agricultura tiene el potencial de poder asistir y conducir a nuevos objetivos (targeted) para su aplicación en programas de producción que van dirigidos al mejoramiento de rasgos de calidad basados en metabolitos (45,50).

3.2.Métodos de Análisis multivariado

3.2.1. Análisis de Componentes Principales (PCA)

El análisis de componentes principales (PCA) es un método de proyección no supervisado que nos permite resumir y visualizar la información de un conjunto de datos de individuos o muestras descritos por múltiples variables cuantitativas correlacionadas. Su objetivo es revelar la estructura interna de los datos que mejor expliquen su varianza o, dicho de otra manera, el PCA trata de encontrar las direcciones con máxima variabilidad (varianza) en los datos. Para ello, este método realiza una transformación ortogonal, T , de un conjunto de datos, X , con variables, en general, correlacionadas a otro conjunto de nuevas variables, conocidas como componentes principales, que no están correlacionadas linealmente. Al realizar este

tipo de transformación, se reduce la dimensionalidad de la data conservando la mayor cantidad posible de información presente en el conjunto de datos. El modelo provee un diagrama donde los componentes score (t) se utilizan para la visualización de la nueva data transformada observándose la relación que existen entre los individuos (Figura 1) (51–53).

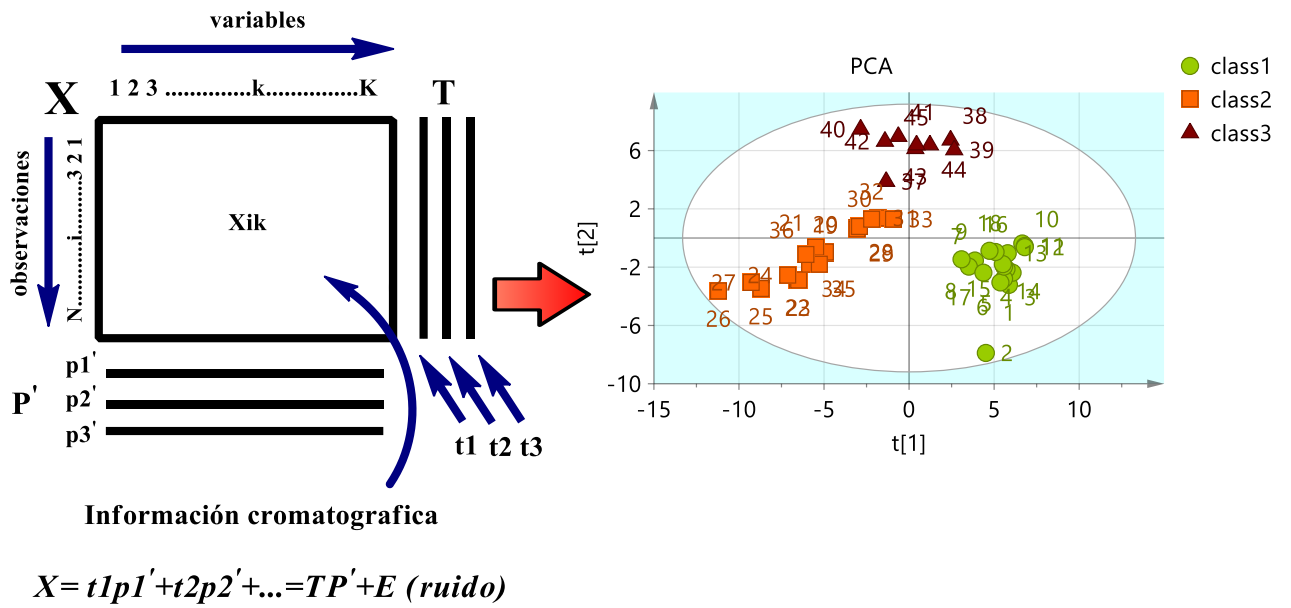


Figura 1. Esquema de transformación de la data cromatográfica por PCA para la reducción de la dimensionalidad (51–53)

3.2.2. Mínimos cuadrados Parciales (PLS)

PLS es un método utilizado comúnmente donde se busca una relación cuantitativa entre dos tablas de matriz de datos X e Y, donde X, usualmente comprende data cromatográfica o espectral de un conjunto de muestras de calibración, y otra matriz Y conteniendo valores cuantitativos, por ejemplo, concentraciones de metabolitos endógenos (Figura 2). PLS es también utilizado en análisis discriminante como PLS-DA. La matriz Y contiene valores cualitativos, por ejemplo, clase pertenencia de clase, género y tratamiento de muestras. El modelo PLS puede ser expresado por:

$$\text{Modelo de X: } X = TP^T + E$$

$$\text{Modelo de Y: } Y = TC^T + F$$

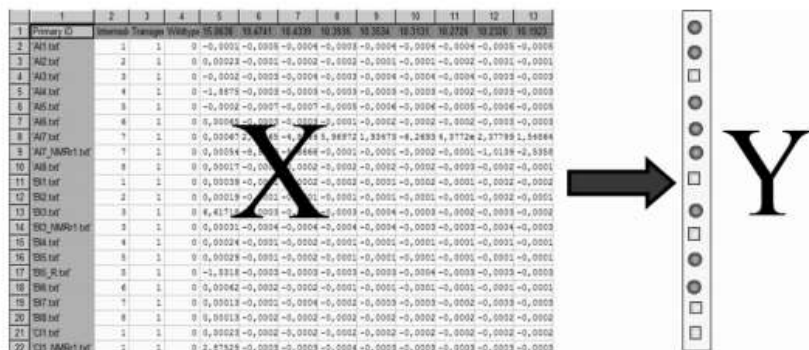


Figura 2. Modelo de construcción de tablas para análisis por PLS

3.2.3. Análisis discriminante de mínimos cuadrados parciales ortogonales

El Análisis discriminante de mínimos cuadrados parciales ortogonales OPLS-DA o también (análisis discriminante de proyecciones ortogonales de estructuras latentes) es un método de proyección supervisado, una modificación de PLS que nos permite hacer una clasificación de los individuos. OPLS separa la variación sistemática de X en 2 partes, una que es linealmente relacionada a Y, y otra que es ortogonal a Y. El modelo OPLS comprende dos bloques de la variación modelada: (1) el bloque predictivo ($T_p P_p'$) que representa la variación entre clases y (2) el bloque ortogonal $T_0 P_0'$ referida a la variación no correlacionada, que representa la variación intra clase (51–53). Esquemáticamente se da en la figura 3.

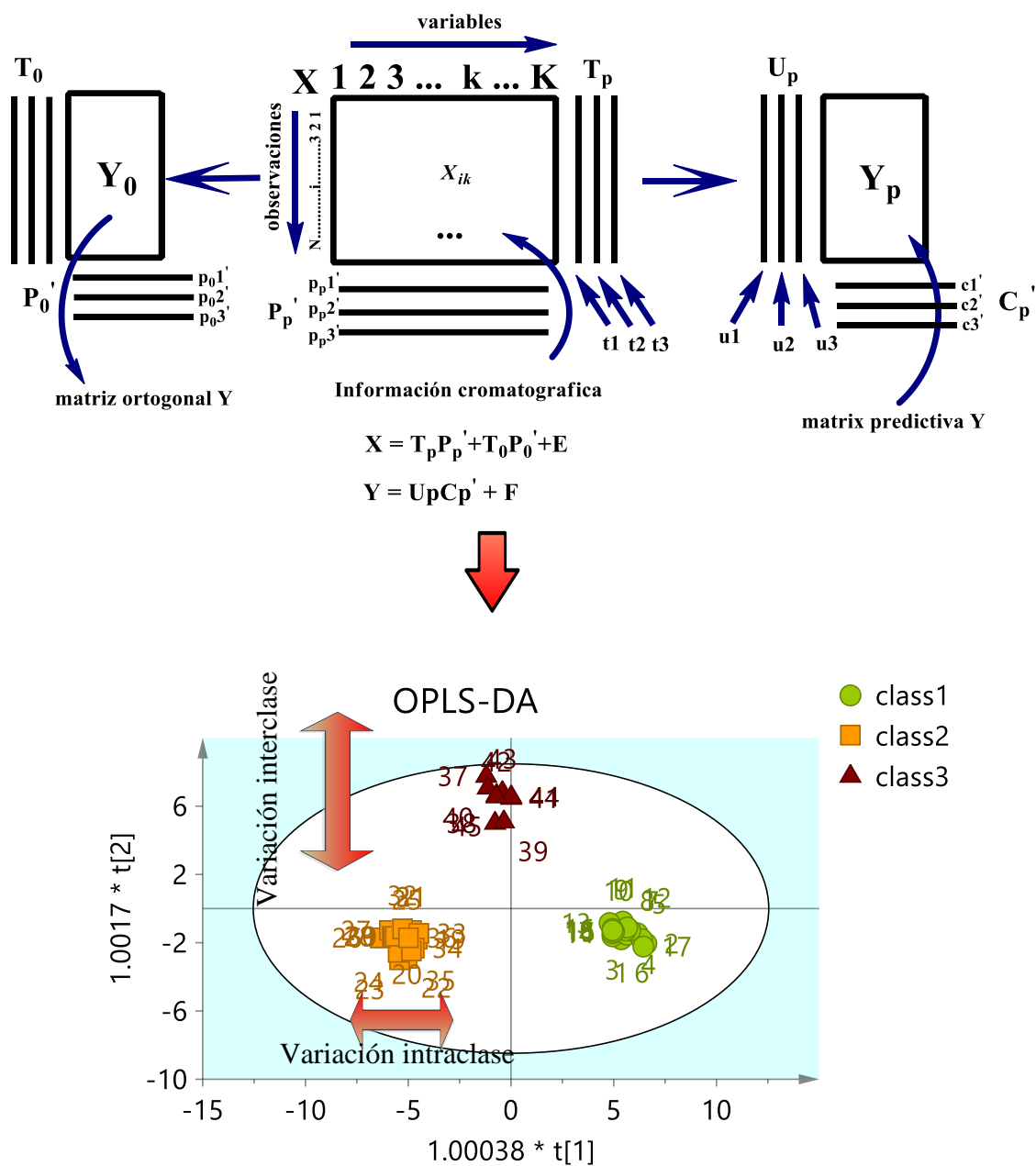


Figura 3. Esquema del algoritmo para el Análisis OPLS (51–53)

El propósito de muchas de las herramientas de trabajo en el campo metabolómico es destacar las diferencias entre muestras bajo diferentes grupos. Las técnicas no supervisadas, tales como PCA destacan la estructura de la data basada solo en la

matriz de mediciones y son un buen punto de partida. Sin embargo, la estructura que muestra puede no estar directamente relacionada con los grupos de interés, ya que las mayores fuentes de varianza pueden no estar relacionadas con ellos. Así, es útil aplicar métodos supervisados que hacen uso de la información del miembro del grupo en el paso de modelado. Por ejemplo, hay una gran cantidad de técnicas semejantes al PCA que realizan una rotación de ejes, pero donde el criterio de elección de los ejes es diferente y utiliza la información del miembro del grupo. El análisis discriminante PLS es una técnica de este tipo (el criterio es maximizar la covarianza entre la matriz de mediciones, y el miembro del grupo codificado), tal como el análisis de variables canónicas (donde el criterio es maximizar la relación de la varianza Inter grupo sobre la varianza entre grupos). Estas técnicas son muy útiles ya que nos pueden permitir resaltar características que no están claras u oscurecidas con los análisis no supervisados. Sin embargo, son propensos a producir resultados que son sobre optimistas debido a que están corridas hacia la selección de características del miembro del grupo para el conjunto de datos, y esta información puede ser o no ruido de fondo (51–53).

3.2.4. Validación cruzada

La razón para el sobreajuste es que la información del grupo es utilizada en el paso del modelado, que puede llevar a circunstancias donde nosotros podemos encontrar un modelo que describe muy bien el modelo de la data, pero este es específico para esa data en particular. La mejor forma para evitar tal situación es asegurar que

donde el modelo aplica la nueva data, ello podría generalizarse. Esta es la propuesta del modelo de validación. En su forma más simple puede llevarse a cabo al dividir el conjunto de datos en dos subconjuntos, un conjunto de entrenamiento (o modelado) y un conjunto de prueba (o validación), preferiblemente con proporciones similares de muestras de los diferentes grupos en ambos subconjuntos. El modelo es entonces construido utilizando el conjunto de entrenamiento y aplicado al conjunto de prueba. Es el resultado de la clasificación del conjunto de prueba que nos informa sobre la habilidad del modelo para clasificar las muestras. Existen posibles variantes para este esquema, más notablemente cuando el tamaño del conjunto de datos es tal que su subdivisión conduce a conjuntos inaceptablemente pequeños. En tales casos, el procedimiento «dejar uno fuera» (también llamado navaja) es una adecuada alternativa; una muestra se toma como conjunto de prueba y clasificada utilizando un modelo construido en todas las otras muestras (Figura 4), y esto se repite, sucesivamente, hasta que todas las muestras han sido tomadas. El rendimiento del modelo se estima luego amalgamando los resultados de todos los conjuntos de prueba **(51–53)**.

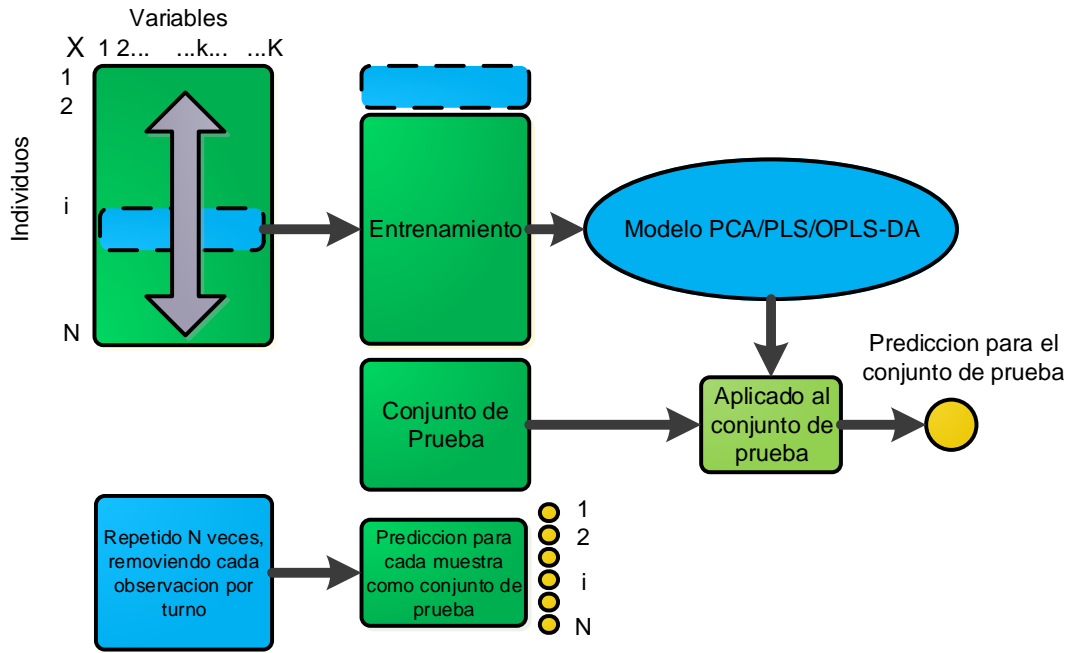


Figura 4. Esquema de procedimiento de validación de «sacar uno fuera». En este y en el procedimiento de dejar un segmento fuera, cada observación sirve una vez como prueba.

3.2.5. R^2 y Q^2

La evaluación de la calidad estadística Q^2 es típicamente reportada como resultado de una validación cruzada y provee una medida cualitativa de consistencia entre la data predicha y original. Incluso aunque Q^2 no tiene un estándar de comparación o valor crítico para inferir significancia, aparte de su máximo valor teórico de 1 o un valor aceptable empíricamente inferido de ≥ 0.7 para un modelo biológico. Desafortunadamente, un modelo inválido o irrelevante es aún capaz de producir un gran valor Q^2 , dado que la validación cruzada consistente requiere de una eliminación sistemática de grandes porciones de data durante su entrenamiento. También, mientras que no haya una estricta medida de validación cruzada, los valores de R^2 de un modelo dado pueden ser usados para evaluar el grado de ajuste de los datos (54).

3.2.6. S-plot

El S-plot es utilizado para visualizar tanto la estructura de la covarianza y la correlación entre las variables-X y el score predictivo t1. Así el S-plot es un diagrama de dispersión de los vectores p1(corr) vs p1 de la componente predictiva. El vector anterior es el vector normal loadings re-expresado como la correlación entre cada variable-X y el vector score t. Con escalado pareto o con centrado, pero no scaling, este plot a menudo toma la forma de la letra S, de ahí su nombre. Las variables-X situadas lejos (extremos) de las alas de la S combinan alta influencia en el modelo con alta confiabilidad y son de relevancia en la búsqueda de variables discriminantes o biomarcadores. La confiabilidad implica una variable-X con una fuerte correlación con t1 y mayor certeza como una variable discriminante o biomarcador putativo. Cabe señalar que el principio de S-plot se aplica a información ortogonal. Cuando la información ortogonal es una parte significativa del modelo, es recomendado examinarlo utilizando S-plot. El S-plot visualiza las variables de influencia en un modelo, este es un diagrama de dispersión que combina los perfiles de loadings de la covarianza y correlación resultantes de un modelo basado en proyección del modelo OPLS-DA (55).

3.3. Métodos de tratamiento de datos

3.3.1. Normalización

Esta es una operación de filas que es aplicada a la data de cada muestra y comprende métodos para hacer la data de todas las muestras directamente comparables con cada otra. Comúnmente se remueve o minimiza los efectos de la variable dilución de las muestras. En conjuntos de datos de MS donde cada sustancia puede tener varios picos, es concebible que el mismo metabolito sea medido más de una vez y, por lo tanto, no hay una razón a priori de que estos datos deban ser diferentes (56).

3.3.2. Escalado

Esta operación se realiza en la columna de los datos (es decir, sobre cada intensidad espectral a través de todas las muestras). Existen un número métodos de escalado que se utilizan muy a menudo (Tabla 1). Cada columna de la tabla puede ser transformada a una media cero al sustraer la media de la columna de cada valor en la columna (centrado en la media). Esto se hace típicamente para que todos los componentes encontrados por el PCA tengan como su origen el centroide de los datos, resultando en un modelo parsimonioso. En segundo lugar, cada columna de la tabla puede escalarse para que tenga una varianza unitaria, al dividir cada valor de la columna por la desviación estándar de la columna. Si la data está centrada a la media, el promedio refleja la covarianza de las variables, mientras que en el escalado de varianza unitaria, el promedio refleja su correlación. Otras formas de escalado son posibles y son muy utilizadas, tales como el escalado Pareto, donde cada variable es dividida por la raíz cuadrada de la desviación estándar de los valores de las columnas. Además, debe quedar claro que las operaciones de

normalización y escalado sirven para diferentes propósitos sin embargo es posible utilizar una combinación de ambos métodos (56,57).

Tabla 1. Lista de los métodos más comunes de escalado utilizados en análisis multivariado

Método	Ecuación	Metas	ventaja	desventaja
Centrado	$\tilde{x}_{ik} = x_{ik} - \bar{x}_k$	Enfocado sobre no diferencias, similaridades	Elimina el desplazamiento de los datos	Inadecuada para data heteroscedastica
UV (varianza unitaria)	$\tilde{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k}$	Compara metabolitos basado sobre su correlación	Todos los metabolitos son igualmente importantes	Aumento de errores de medida
Pareto	$\tilde{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sqrt{s_k}}$	Reduce la relativa importancia de grandes valores, parcialmente preserva la estructura de la data	Mantiene más cerca la medida original que UV	Sensible a factores de cambio

3.4. Análisis factorial múltiple (AFM)

El análisis factorial múltiple permite analizar simultáneamente varias tablas de variables y obtener resultados en particular representaciones gráficas que permiten estudiar la relación entre las observaciones, las variables y las tablas. Al interior de la tabla las variables deben tener la misma naturaleza (cuantitativas o cualitativas), sin embargo, las tablas pueden ser de diferentes tipos. El AFM es un resumen de PCA (análisis de componentes principales) y ACM (análisis de correspondencia múltiple) que permite homogenizar las variables heterogéneas. La metodología del AFM comprende dos etapas:

1) Se realiza sucesivamente por cada tabla un PCA o un ACM en función de la naturaleza de las variables. Se conserva el valor del primer valor propio de cada uno

de los análisis al ponderar enseguida las diferentes tablas en la segunda parte del análisis.

2) Se realiza a continuación un PCA ponderado sobre las columnas del conjunto de tablas, sabiendo que las tablas de variables cualitativas son transformadas en tablas disyuntivas completas. Cada una de las indicatrices de las tablas disyuntivas tienen un peso en función de la frecuencia de la modalidad correspondiente. La figura 5 muestra cómo es la disposición de las tablas de datos para un análisis factorial múltiple (58).

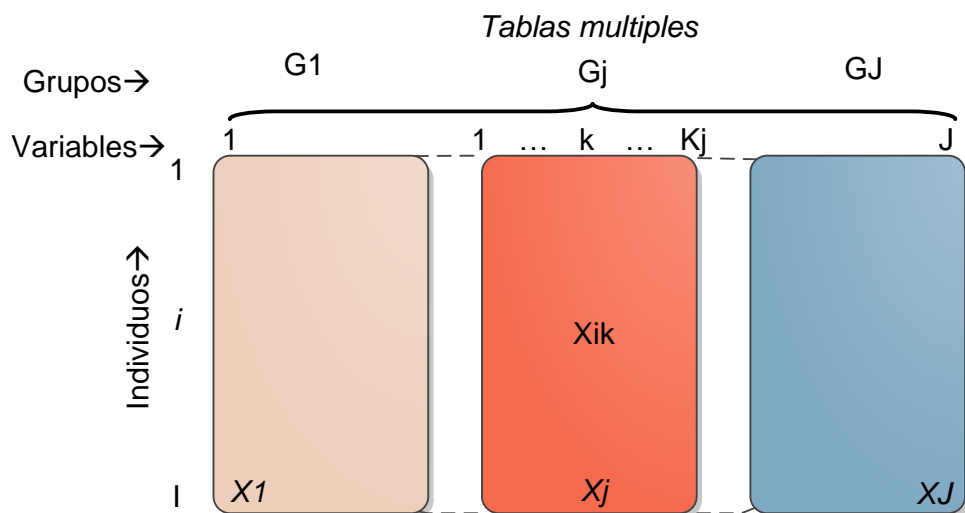


Figura 5. Esquema general de un AFM constituida por j tablas cuantitativas (tablas centrales), tablas cualitativas de la izquierda y tablas ilustrativas de la derecha

3.4.1. Objetivos de un AFM

En un AFM se estudian las semejanzas entre individuos desde el punto de vista del conjunto de variables y las relaciones entre las variables. Se toma en cuenta la estructura en grupos aprovechando la estructura del grupo y de las variables, es decir se estudia globalmente las semejanzas y las diferencias entre los grupos

observando la especificidad de cada grupo, si la variable de un grupo aporta la misma información de la variable de otro grupo o bien si existe una información común o una información específica aportada por el grupo o bien aún no existe ninguna vinculación entre la información aportada por ese grupo y que no es aportado por el otro grupo. Para hacer esto uno puede observar un individuo particular descrito de la misma manera para todos los grupos o bien si hay un grupo que los describe de una manera muy específica. Al final uno podrá comparar la tipología obtenida por cada grupo (comparar tipologías a partir de análisis separados). Un punto crucial del análisis multitaslas es el equilibrio de la información. La información de un grupo no debe ser aplastada u ocultada por la información aportada por los otros grupos y por lo tanto se debe equilibrar la información aportada por cada grupo.

3.5.Diagramas Heatmap

El perfil metabolómico puede ser visualizado utilizando un Heatmap. Esta imagen colorizada de la matriz de datos puede ayudar efectivamente a comunicar claras diferencias en la data debido a que las variables (y observaciones) son reordenadas a estructuras destacadas en la data. Esta es una de las formas de visualizar el agrupamiento jerárquico de clústeres. El primer agrupamiento jerárquico se realiza tanto en las filas como en las columnas de la matriz de datos. Las filas /columnas de la matriz de datos se reordenan de acuerdo con el resultado del agrupamiento jerárquico poniendo las observaciones similares una cerca de la otra. Los bloques

con valores "alto" y "bajo" valor son adyacentes en la matriz de datos. También se le llama imagen en falso color, donde los valores de los datos son transformados a escala de colores. La estructura en la data es representada tanto en bloques de colores, así como dendogramas a la izquierda y en la parte superior. Para realizar un Heatmap típicamente están involucradas tres pasos a seguir: primero, se debe definir una distancia de medida; segundo, se debe escoger un algoritmo de agrupamiento (clustering) y tercero, se debe construir un esquema de color. Los Heatmaps son utilizados comúnmente para clustering no supervisados que son construidos basados sobre los potenciales candidatos de importancia o VIP (variables de importancia), que son extraídos a partir de un análisis por OPLS-DA. A primera vista, las réplicas de las muestras (columnas) deben ser agrupadas juntas. Luego hay generalmente algunos bloques de grupos de color de metabolitos que muestran similares patrones de expresión a través de múltiples muestras. Finalmente, se aplica un esquema de color para la visualización y se muestra la matriz de datos. Visualizar la matriz de datos de esta manera puede ayudarnos a encontrar las variables que parecen ser características para cada clúster de muestra (59–61).

3.6.CV-ANOVA

El diagnóstico CV-ANOVA se basa en la CV (validación cruzada) para la estimación de predictores independientes (scores OPLS) y residuos predictivos. En el ANOVA se compara dos modelos ajustados a los mismos datos por el tamaño de

sus residuos ajustados. En el contexto de regresión, los dos modelos que se comparan son:

$$y_i = \text{constante} + d_i \dots (1)$$

$$y_i = \text{constante} + bx_i + e_i \dots (2)$$

El ANOVA se realiza sobre el tamaño de la suma de los cuadrados (SS (d)) y (SS (e)), estos no son independientes ya que la respuesta (y) es la misma. En otras palabras, se evalúa si el modelo PLS ó OPLS tiene residuos predictivos de CV significativamente más pequeños que solo la variación alrededor del promedio global.

También, CV-ANOVA utiliza una prueba F para la prueba de significancia (prueba de hipótesis) de la hipótesis nula de residuos iguales de los dos modelos. Las pruebas F suponen que los residuos de los dos modelos comparados están aproximadamente distribuidos normalmente. El resultado de CV-ANOVA utiliza el diseño convencional ANOVA **(62)**.

En la tabla ANOVA, se muestran los siguientes resultados (tabla 2):

Tabla 2. Significado de los resultados de CV-ANOVA

Vector	Tipo	Descripción
SS	Total correguido	SS del Y del conjunto de trabajo (workset) correguido por la media.
	Regresión	Fracción del Total Correguido SS explicado por el modelo, estimado a través del principio de validación cruzada.
	Residual	Diferencia entre total correguido y regresión SS, es decir, la fracción del total correguido no contabilizado por el modelo.
DF	Total correguido, Regresión, Residual	El número de grados de libertad. (DF). Este es un número aproximado basado en la experiencia de que PLS necesita la mitad de los componentes para alcanzar la misma explicación de Y que la regresión de componentes principales.
MS	Total correguido, Regresión, residual	Al dividir cada SS por el DF respectivo, se obtienen los cuadrados medios correspondientes (MS), o varianzas.
F		La prueba F, basada en la relación Regresión de MS / Residual de MS, evalúa formalmente la importancia del modelo.
<i>p</i>		El valor <i>p</i> indica el nivel de probabilidad donde un modelo con este valor F puede ser el resultado de una casualidad. La práctica común es interpretar un valor <i>p</i> inferior a 0.05 señalado como un modelo significativo.
SD	Desviación estándar	Raiz cuadrada de MS.

3.7. Curvas ROC (Característica Operativa del Receptor)

Una curva ROC es un resumen gráfico del rendimiento de un clasificador binario. El análisis ROC utiliza el área bajo la curva AUC como criterio para evaluar la discriminación entre dos clases. El valor AUC es una medida directa del rendimiento de un clasificador que representa qué tan bien están separadas dos clases. La capacidad de identificar correctamente un evento como positivo se llama sensibilidad, también conocida como Tasa Positiva Verdadera (TPR). Inversamente, identificar correctamente los eventos negativos se llama especificidad o Tasa Negativa Verdadera (TNR). El riesgo de identificar positivamente un evento negativo es la tasa de falsos positivos (FPR) y el riesgo de identificar negativamente un evento positivo es la tasa de falsos negativos (FNR). Las relaciones matemáticas que conectan estos cuatro criterios se dan en las

ecuaciones (1, 2 y 3). El análisis ROC trata por separado la capacidad de identificar un evento positivo como positivo y un evento negativo como negativo. Al considerar dos vectores de respuestas continuas cuyos valores son 0 y 1, cada vector representaría un grupo (positivo o negativo) que debe ser discriminado del otro y cuanto mejor sea la separación entre las dos distribuciones, mayor será el valor de AUC (Figura 6) (63,64).

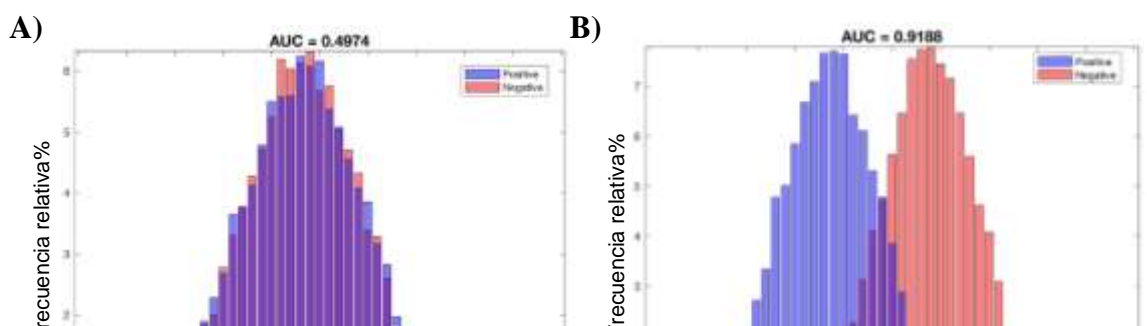
Definición de términos utilizados en las curvas ROC y AUC:

$$\text{TPR (sensibilidad)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots (1)$$

$$\text{especificidad} = \frac{\text{TN}}{\text{TN} + \text{FP}} \dots (2)$$

$$\text{FPR} = 1 - \text{especificidad} = \frac{\text{FP}}{\text{TN} + \text{FP}} \dots (3)$$

La curva ROC es la representación de la TPR registrada en función de la FPR (Figura 7). En SIMCA, se puede crear una gráfica de una curva ROC para modelos de clase y análisis discriminante (DA). El AUC representa qué tan bien están separadas dos clases, y también puede representar un valor global de qué tan bien están separadas varias clases.



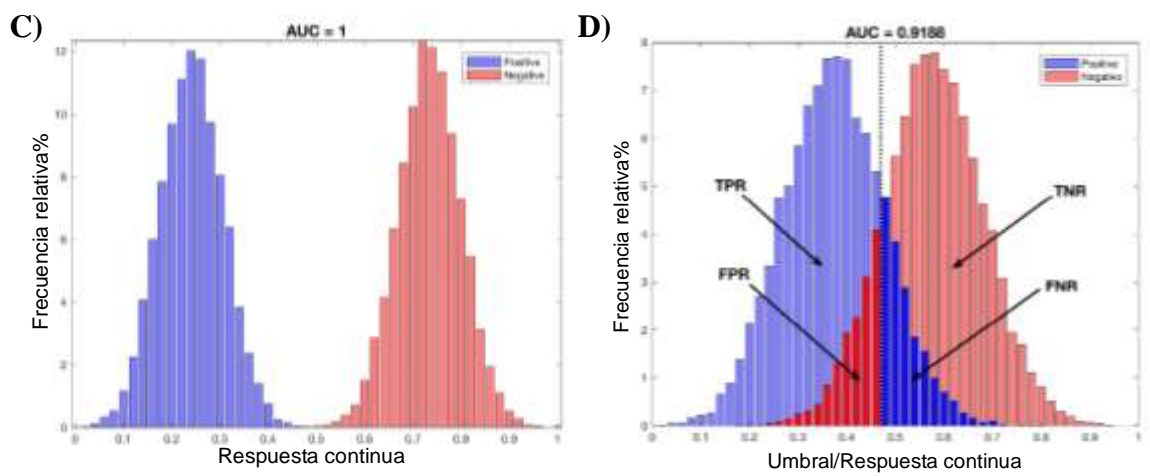


Figura 6. Distribuciones de respuesta para muestras no relacionadas (A), muestras relacionadas (B), muestras bien separadas (C). En (D) se muestran el gráfico de muestras relacionadas (positivas) y muestras no relacionadas (negativas) que muestran un umbral posible entre 0 y 1 maximizando TPR / TNR y minimizando FPR / FNR (65).

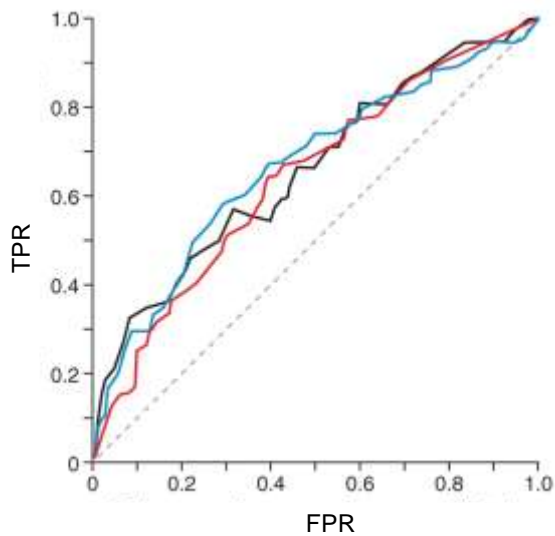


Figura 7. Curva típica ROC

3.8. Combinación de tablas de datos ESI (+, -)

Los estudios non-targeted requieren generalmente del análisis por duplicado de todas las muestras una con cada modo de ionización para incrementar la cobertura de metabolitos. El análisis por duplicado del conjunto total de las muestras produce dos conjuntos de data independientes que pueden contener información de entrada correspondiente a metabolitos comunes. Algunos metabolitos tales como aminoácidos, ácidos carboxílicos de cadena corta, purinas, etc. pueden ser detectados en ambos modos de ionización e imposibilita la combinación de los dos conjuntos de datos en una simple matriz debido a que el conjunto de datos resultante contendrá información redundante sobre aquellos metabolitos detectados en ambos modos. Como resultado el análisis de la data frecuentemente requiere utilizar diagramas de flujo independientes para procesar los dos conjuntos de datos (ionización positiva y negativa) y la utilización por el usuario de un extenso tiempo

de procesamiento. También, la data puede solamente ser interpretada después de que los metabolitos significativos detectados en cada modo de ionización se hayan identificado de forma independiente. Los últimos avances en tecnología de polaridad de conmutativa rápida y robusta han permitido la adquisición simultánea de data de ionización positiva y negativa en un solo análisis, lo que acorta los tiempos de procesamiento. Sin embargo, este enfoque ha mostrado ser ineficiente en el análisis untargeted ya que disminuye la resolución y la sensibilidad a través de una combinación de ciclos para MS scan y fragmentación MS /MS en ambas polaridades.

La identificación de las características moleculares detectadas en ambos modos de ionización antes del análisis estadístico permitiría utilizar un único conjunto de datos que consistiría en una única base de datos de metabolitos. El Package MSCombine permite el análisis estadístico de un único conjunto de datos por lo que evita la necesidad del análisis en paralelo uno por polaridad. El software R se utiliza para configurar el algoritmo de MSCombine, la figura 8 muestra los pasos para la combinación de ambas tablas de datos en ESI (+) y ESI (-).

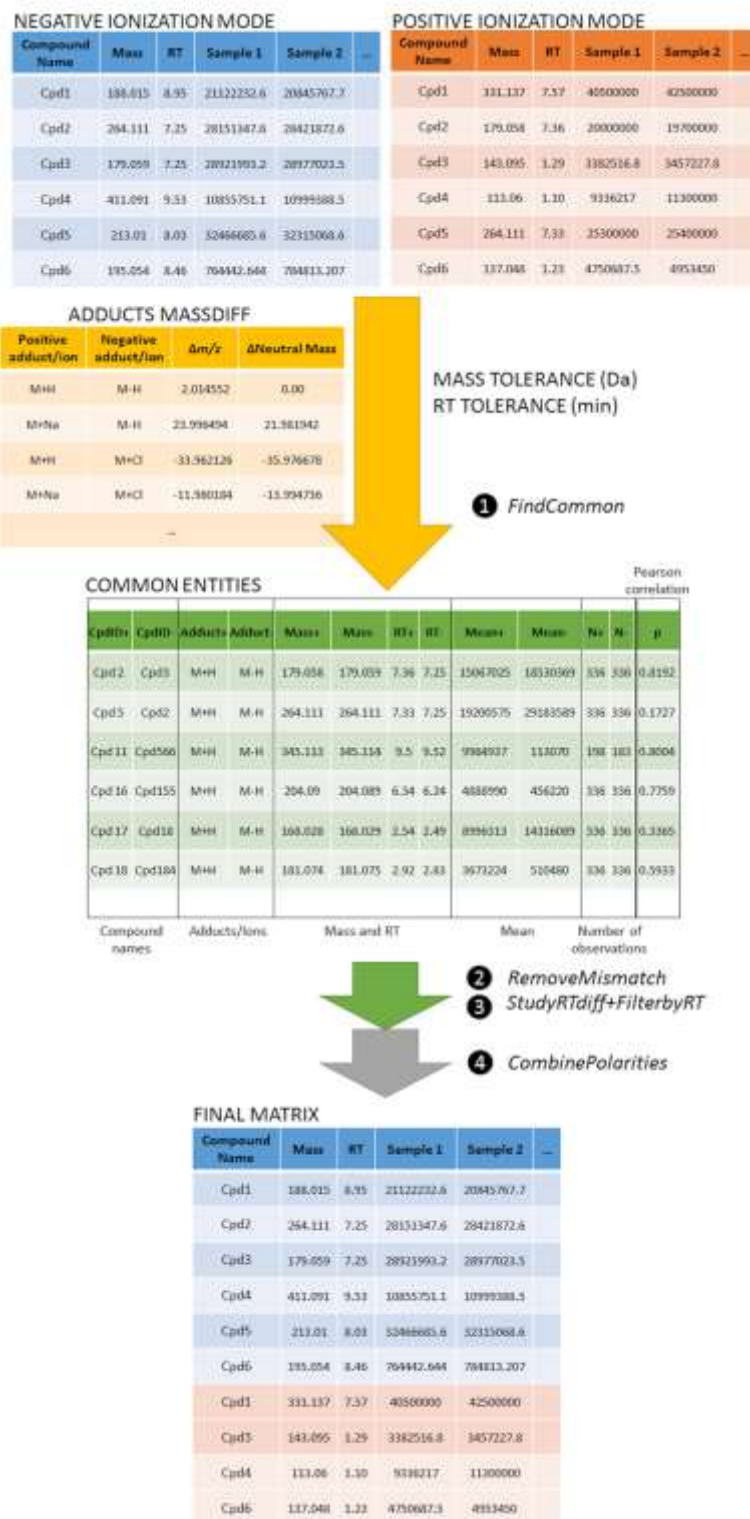


Figura 8. Arquitectura de MScombine, que utiliza las siguientes funciones: (1) FindCommon, (2) RemoveMismatch, (3) StudyRTdiff y FilterbyRT, y (4) CombinePolarities (66).

3.9. Análisis multibloque

El análisis multibloque es un reciente método de reducción de dimensionalidad de data multivariada y tiene como objetivo la integración de múltiples tablas de datos ómicas (Package de R MixOmics). Este análisis permite clasificar e identificar una huella molecular multi-ómica. El análisis multibloque conocido como DIABLO (Data Integration Analysis for Biomarker discovery using a Latent Component method for Omics studies), modela la estructura de correlación entre las tablas de datos ómicas dando como resultado una capacidad mejorada para asociar biomarcadores a través de múltiples niveles funcionales en fenotipos de interés. Los resultados gráficos de este método ayudan a la interpretación de análisis complejos y pueden proporcionar conocimientos biológicos significativos. Este método fue particularmente útil en aplicaciones en química ecológica ya que esta ciencia es fundamentalmente una ciencia comparativa (67). En la figura 9 se muestra el esquema del método de Integración de múltiples tablas de datos Ómicas.

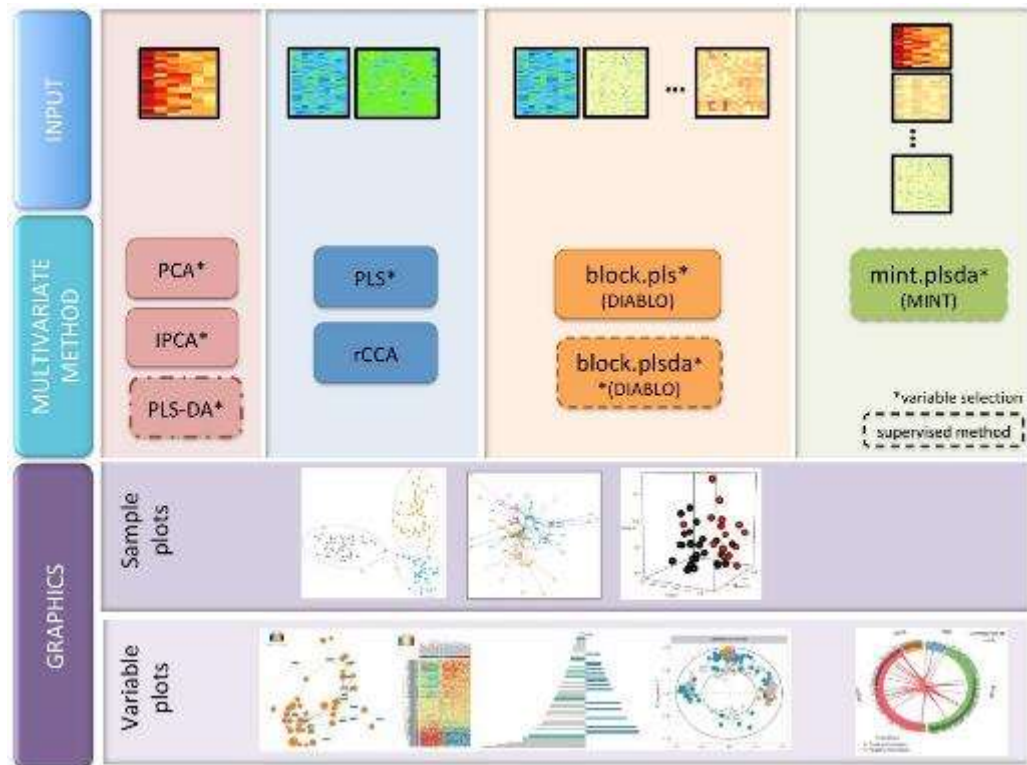


Figura 9. Visión general de los métodos multivariados de MixOmics para análisis supervisados para análisis de una simple e integración de tablas de datos Ómicas (68).

3.10.Métodos de análisis Univariado

3.10.1. Análisis de varianza (ANOVA)

Un ANOVA prueba la hipótesis de que las medias de dos o más poblaciones son iguales. Los ANOVA evalúan la importancia de uno o más factores al comparar las medias de las variables respuesta en los diferentes niveles de los factores. La hipótesis nula establece que todas las medias de la población (medias de los niveles de los factores) son iguales mientras que la hipótesis alterna establece que al menos una es diferente. Para realizar un ANOVA se debe tomar en cuenta los siguientes supuestos: 1) la variable independiente debe medirse al menos a nivel de intervalo 2) independencia de las observaciones, 3) la distribución de los residuales debe ser normal 4) Homocedasticidad, las varianzas deben ser homogéneas.

3.10.2. Análisis Post-hoc, Test de Tukey

El ANOVA permite ver diferencias significativas entre los grupos, pero no muestra si existen diferencias significativas entre los grupos tomados dos a dos. El test de Tukey permite realizar múltiples comparaciones entre las medias. El test se realiza para analizar las diferencias entre las medias con un intervalo de confianza al 95%.

3.10.3. Perfil sensorial

La presentación gráfica de los datos sensoriales implica el uso de gráficos de radar, la superposición de los gráficos de radar se utiliza para comparar los atributos de los alimentos entre sí al graficar las intensidades relativas (69,70). Un diagrama de barras ofrece la misma información que un gráfico de radar, adicionalmente un ANOVA permite evaluar las diferencias significativas entre los atributos para una mejor descripción del alimento en estudio (71).

3.11. Bases de datos de Productos Naturales

3.11.1. KNApSAcK

El corazón de la base de datos de KNApSAcK contiene información de las relaciones entre los metabolitos y su organismo de expresión con un total de 101,500 relaciones especies-metabolitos que abarcan 20,741 especies y 50,048 metabolitos. Su página web contiene un motor de búsqueda para el uso en investigación metabolómica, lo que hace posible buscar metabolitos basados en su masa exacta, fórmula molecular, el nombre del metabolito o espectros de masas en varios modos de ionización (72,73).

3.11.2. Northern African Natural Products Database (NANPDB)

La NANPDB Recopila la información de 4500 productos naturales. La base de datos cubre algunos compuestos aislados principalmente de plantas y

contribuciones de algunos endófitos, animales (por ejemplo, coral), hongos y fuentes de bacterias. Los compuestos fueron identificados a partir de 617 especies bajo 146 familias. Las propiedades fisicoquímicas computadas a menudo utilizadas para predecir el metabolismo y la farmacocinética de la droga, así como predecir la información de toxicidad ha sido incluida para cada compuesto en la base de datos. Esta es una gran colección de compuestos naturales anotados y producidos por organismos del norte de África (74).

3.11.3. Plant Metabolic Network (PMN, PlantCyc)

PlantCyc es una base de datos de referencia de rutas metabólicas que contiene más de 800 rutas con sus enzimas y genes catalíticos, así como compuestos de más de 350 especies de plantas. Las rutas metabólicas de un gran número de plantas están catalogadas de manera integral en la base de datos PlantCyc de PMN (75).

3.11.4. Universal Natural Products Database (UNPD)

Es la mayor base de datos no comercial y de libre acceso para productos naturales, comprende 197201 productos naturales de plantas, animales y microorganismos. El UNPD contiene un buen número de moléculas solo publicadas en publicaciones chinas o incluso algunas de ellas en la actualidad no han sido aún publicadas (76). La reciente publicación de la Base de Datos Universal de Productos Naturales (UNPD), hace una compilación de 229 358 productos naturales entre macro y

microorganismos terrestres y marinos que se ha puesto a disposición del público (77).

3.11.5. Chemical Entities of Biological Interest (ChEBI)

ChEBI es una base de datos para entidades químicas de bajo peso molecular que son relevantes para comprender el funcionamiento biológico. En los últimos años, ChEBI ha crecido constantemente en contenido anotado, en la publicación de noviembre de 2012 contiene cerca de 30 000 entidades totalmente anotadas Toda la data de ChEBI está disponible en PubChem (78).

3.11.6. LipidMAPS

La base de datos de estructura de LIPID MAPS (LMSD) es una base de datos relacional que abarca estructuras y anotaciones de lípidos biológicamente relevantes. A partir del 18 de octubre del 2018, LMSD contiene 43168 estructuras lipídicas únicas por lo que es la mayor base de datos pública de lípidos del mundo (79–81).

3.11.7. Dictionary of Natural Products (DNP)

El Diccionario de Productos Naturales en DVD brinda un rápido acceso a datos químicos, físicos y biológicos de más de 259,859 compuestos contenidos en más de 68,000 entradas. Revisada continuamente para mantener el ritmo de la literatura

actual, junto con la inclusión de las bases de datos Berdy Bioactive Natural Products y DEREPE, esta estructura de base de datos es la fuente más completa de información sobre productos naturales disponible **(82,83)**.

3.11.8. NIST 14

La base de datos NIST 14 es una biblioteca de espectros de masas de ionización electrónica (EI) con 220,460 espectros revisados por pares de 192,108 compuestos únicos, con identificaciones y generalmente estructuras químicas. Contiene una biblioteca de datos de cromatografía de gases (GC) con 224,038 valores de índice de retención de Kovats para 21,847 compuestos en la biblioteca EI, tanto en columnas polares como no polares e incluye condiciones de columna GC y citas bibliográficas. NIST08 Mass Spectral Library contiene espectros, estructuras y citas bibliográficas de: medicamentos, metabolitos y venenos; pesticidas y fungicidas; orgánicos presentes en el suelo, el agua y el aire; aminoácidos, di- y tripéptidos; muestras comunes contaminantes y derivados analíticos comunes de los anteriores **(84,85)**.

3.11.9. Global Natural Products Social Molecular Networking (GNPS)

La base de datos GNPS es una plataforma basada en datos para el almacenamiento, el análisis y la difusión del conocimiento de los espectros de masa MS/MS que permitan el intercambio comunitario de espectros de masa MS² sin procesar, la anotación continua de espectros depositados y la curación colaborativa de espectros

de referencia (referidos a librerías espectrales) y espectros de masas experimentales (organizados como conjuntos de datos).

GNPS permite analizar espectros de masa MS/MS y compararlos con todos los espectros de masa disponibles públicamente. En la actualidad, GNPS utiliza 221083 espectros MS/MS de 18163 compuestos únicos para las búsquedas. Estos incluyen bibliotecas de terceros, tales como MassBank, ReSpect y NIST, así como bibliotecas espectrales creadas para GNPS (GNPS-Collections) y espectros de la comunidad NP (GNPS-Community) (86).

3.12. Ajíes Nativos Peruanos (*Capsicum* spp.)

Las plantas del género *Capsicum* perteneciente a la familia botánica de las Solanaceae junto con otras plantas tales como el tomate, berenjena, papa, o tabaco, son algunas de las plantas más antiguas cultivadas en el mundo. Sobre la base de la clasificación taxonómica del género *Capsicum* se incluyen al menos 38 especies de las que 5 especies son domesticadas y económicamente más notables; estas son: *Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, *Capsicum baccatum*, y *Capsicum pubescens* (9).

Debido a sus características de pungencia, aromas, y sabores, los frutos de los ajíes son un importante ingrediente para millones de personas en la dieta diaria. Además, son una buena fuente de vitaminas antioxidantes C y E y provitamina A, así como una fuente de otros antioxidantes que evitan la oxidación de lípidos vía captura de

radicales libres. También se han analizado como protectores contra el cáncer, anemia, diabetes, y enfermedades cardiovasculares (87).

3.13. Taxonomía del género *Capsicum*

Las cinco especies están agrupadas en un complejo taxonómico el cual tiene convencionalmente tres, o tal vez 2 o más especies, con tres grupos de plantas domesticadas que parecen ser más divergentes que sus progenitores salvajes. Las dos especies domesticadas restantes se encuentran en otros complejos taxonómicos del género. Estas dos especies son poco utilizadas fuera de América Latina, mientras que la variedad *C. baccatum* ha sido ampliamente domesticada y hasta la actualidad muy utilizada. Durante algún tiempo, se aplicó el nombre de *C. frutescens* a los ajíes domesticados de *C. annuum*, por lo que en algunas publicaciones se necesita precisar si las plantas que se analizan son en realidad *C. annuum* (que es más probable), o *C. frutescens* o tal vez otra de estas especies (88,89).

Si la morfología es la base primaria para el reconocimiento taxonómico de las especies, entonces es de esperarse que cada una de las 5 especies domesticadas deberían mantener distintas categorías. Sin embargo, la separación morfológica de los taxones espontáneos silvestres aún no está muy esclarecida. Las especies silvestres de *C. baccatum*, *C. frutescens* y *C. chinense* se fusionan morfológicamente para formar fenotipos indistinguibles en el nivel más primitivo.

Además, cada uno de los fenotipos silvestres muestra una serie de formas transitorias de los taxones silvestres a los domesticados. Si se sigue la lógica taxonómica, se debería mantener la variedad *C. annum* var. *glabriusculum* o su equivalente, mientras que se designan dos nuevas variedades para representar las especies silvestres ancestrales de *C. frutescens* y *C. chinense*. Jensen *et al.* (1979) utilizando data de isoenzimas y Nei's (1972) utilizando distancia genética estándar mostraron que la distinción entre *C. annum*, *C. frutescens* y *C. chinense* como especies es a menudo arbitraria. En este análisis, se utilizaron los alelos de varios sistemas enzimáticos como caracteres taxonómicos, mientras que cada muestra fue tratada como una unidad taxonómica operacional (UTO) separada. Esto demostró que dentro de cada especie muchos genotipos son repetidos. El análisis de clúster final de estas UTOs representativas mostraron la completa dispersión de los genotipos *C. annum*, *C. frutescens*, y *C. chinense*. El dendograma y análisis de clúster desarrollado por esta técnica indicó que las tres especies domesticadas no pueden ser distinguidas basados sobre perfiles enzimáticos. El dilema está en que estos dos análisis numéricos independientes sugieren conclusiones muy diferentes. Por un lado, Pickersgill *et al.* encontraron que *C. annum*, *C. frutescens* y *C. chinense* son distintos desde el punto de vista morfológico mientras que Jensen *et al.* encontraron las tres taxas indistinguibles basados sobre perfiles enzimáticos. El problema planteado por Pickersgill y Jensen ha sido la base de la confusión taxonómica dentro de este complejo taxonómico por muchos años. Aunque esta discusión no ha servido para proveer una solución al problema de cómo tratar taxonómicamente los *Capsicum* domesticados, bien puede explicar parte del dilema

que enfrentaron los primeros taxonomistas al no reconocer las diferencias entre las especies taxonómicas. Si esto se considera a la luz de otras variables tales como: variedades, cultivares, razas, formas, etc., creados por el hombre, el dilema para desarrollar una taxonomía racional para el fitomejorador y el horticultor es enorme (90).

3.14. Ajíes del Perú, accesiones de la colección del INIA y UNALM

El Instituto Nacional de Innovación Agraria (INIA) es un instituto dedicado a la conservación y uso de los recursos genéticos en el Perú. En el año 2013, el INIA en colaboración con Bioversity International lograron rescatar 413 accesiones de ajíes provenientes de las colectas en 11 departamentos del Perú, gracias al proyecto “Rescate y promoción de los ajíes nativos en su centro de origen” (91). Los ajíes colectados por el INIA fueron conservados en su banco de semillas. Posteriormente, la colección completa de semillas de ajíes fue sembrada en la Estación experimental del INIA ubicado en Donoso, Huaral, para su multiplicación y caracterización agromorfológica. El estudio taxonómico rindió resultados poco favorables ya que no se pudo determinar para todas las accesiones a qué especie comercial correspondían comparando la taxonomía y utilizando descripciones existentes de los fenotipos conocidos de ajíes comerciales (92).

Por otro lado, el «Proyecto *Capsicum*» del Programa de Hortalizas de la Universidad Nacional Agraria La Molina (UNALM) cuenta con una colección de

accesiones que serían las más representativas de la diversidad peruana, conformada por cuatro especies cultivadas (*C. annum*, *C. baccatum*, *C. chinense*, y *C. frutescens*) provenientes de 21 departamentos del Perú (Figura 10). El proyecto ha logrado realizar avances en caracterización morfológica y molecular de ajíes nativos. El banco de germoplasma de *Capsicum* de la Universidad Nacional Agraria La Molina cuenta con 369 accesiones de ajíes nativos en su mayoría provenientes de parcelas de pequeños agricultores de la costa y selva del Perú. Actualmente la UNALM trabaja con el INIA para publicar el primer catálogo de ajíes nativos del Perú en el marco del programa nacional de innovación agraria (PNIA), el título del proyecto es: « Consolidación de las colecciones INIA y UNALM de *Capsicum* nativos del Peru en una base de datos morfológica, genética, físico-química y culinaria integrada y análisis de las prioridades de conservación y desarrollo » (43,93).



Figura 10. Biodiversidad de ajíes Peruanos (93)

En el año 2016, la Universidad Peruana Cayetano Heredia en colaboración con el INIA, caracterizaron accesiones del banco de semillas del INIA a nivel bioquímico y agromorfológico de gran demanda a nivel local, logrando identificar accesiones con características de interés comercial, con miras a poner en valor los ajíes nativos. Se seleccionaron 50 accesiones promisorias en base a dichas características. Un segundo estudio que duró 2 años permitió evaluar con más detalle las características agromorfológicas, químico-nutricionales, antioxidantes y sensoriales de estas 50 accesiones (Proyecto FINCyT) (94). De este último estudio se realizó la selección de 5 accesiones promisorias por poseer las mejores características. La selección de estas accesiones estuvo a cargo de la Empresa Agroexport Topara productora y exportadora de ajíes y especialistas en gastronomía (Chefs). Los nombres comunes de estas accesiones son: Ayuclo (*Capsicum baccatum*), Miscucho (*C. chinense*), Tomatito rojo (*C. baccatum*), Charapita amarillo (*C. frutescens*) y Dulce rojo (*C. chinense*), cuyos centros de origen son los departamentos de San Martín (Ayuclo y Charapita), Ucayali (tomatito rojo y dulce rojo), Costa de Lambayeque (Miscucho). En una tercera etapa, estas accesiones fueron sembradas y cosechadas en tres localidades de la costa peruana que podrían ser consideradas con un alto potencial para la producción de ajíes, como parte del programa que tiene el INIA de “evaluar el comportamiento y manejo pos cosecha de cultivares promisorios de ajíes nativos bajo las condiciones agroecológicas de la costa peruana con miras a establecer su cultivo y aprovisionamiento sostenible para el sector gastronómico”. Las tres localidades son las siguientes: 1) instalaciones de la Empresa Agroexport Topara ubicada en Chíncha, 2) en la instalación experimental del INIA ubicado en

Chiclayo y 3) en la instalación experimental del INIA Donoso ubicado en Huaral. La producción de los ajíes de Chíncha tuvo un manejo agrícola orgánico mientras que para las otras 2 localidades tuvo un manejo agrícola convencional. Las 5 accesiones de ajíes nativos mencionadas, cultivadas en 3 diferentes localidades, son el objeto de estudio del presente proyecto de investigación.

3.15.El gusto, sabor y aroma

El «sabor» es una combinación de dos sensaciones diferentes, el gusto y el olor (95). Los seres humanos comparten el mismo conjunto de sentidos empleados para interpretar el gusto, el aroma, el sabor y la textura de los alimentos. Al considerar el sentido del gusto, nos referimos a la lengua como el órgano sensorial en la que se perciben las sensaciones. Por otro lado, los receptores químicos que comprenden el gusto, el olfato y la quemestesis son los más relevantes. La quemestesis es la capacidad sensorial de la piel de la nariz y de las mucosas nasales a diversos compuestos. Estos compuestos pueden ofrecer tonalidades que a veces se les denominan de forma popular sabores, un ejemplo es el denominado «*sabor picante*» que técnicamente se denomina como una *quemestesis*. Entre las quemestesis se encuentran sustancias irritantes como la capsaicina, pero pueden encontrarse otras que produzcan sensaciones refrescantes como el mentol. Estos tres sistemas únicos y distintos facilitan la transmisión de información respecto a la presencia de químicos en el entorno inmediato. La quemestesis surge cuando ciertos compuestos químicos activan receptores normalmente asociados con otros sentidos

que median el dolor, el tacto y la percepción térmica, es decir, falso calor y falso frío. El Gusto (degustación) detecta compuestos químicos utilizando sensores principalmente de la boca. El Olor (aroma u olfatación) detecta sustancias químicas en el aire, tanto externamente a través del pasaje nasal como de aquellos compuestos emitidos por los alimentos masticados en nuestra cavidad oral, vía retronasal. El pasaje retronasal es el camino de la percepción olfativa, que es tan importante o más que el pasaje nasal. Por encima de esto, también existen las quemestesis que involucran las sensaciones como la astringencia y la pungencia (Figura 11). Hay que destacar que, para disfrutar del sentido holístico de la comida, el ser humano posee un sentido general del aroma/gusto de olor y sabor. También, el ser humano posee alrededor de 9000 papilas gustativas y varios millones de receptores neuronales en el epitelio olfativo. Los complejos procesos de olfatación y degustación comienzan en el momento en que las moléculas se desprenden de los alimentos y llegan a los sentidos del gusto y olfato y la diferenciación entre miles de gustos y olores es el resultado de procesos complejos que activan varias combinaciones de receptores de gusto y olor. También, las vías neurales de los estímulos del aroma y sabor son diferentes, pero generalmente los dos trabajan al unísono. Asimismo, mucho de lo que los seres humanos perciben como sabor tiene mucho que ver con los aromas de los alimentos. Por último, es necesario mencionar que gran parte de lo que comemos se conecta a las áreas del cerebro responsables de las emociones (96–98).

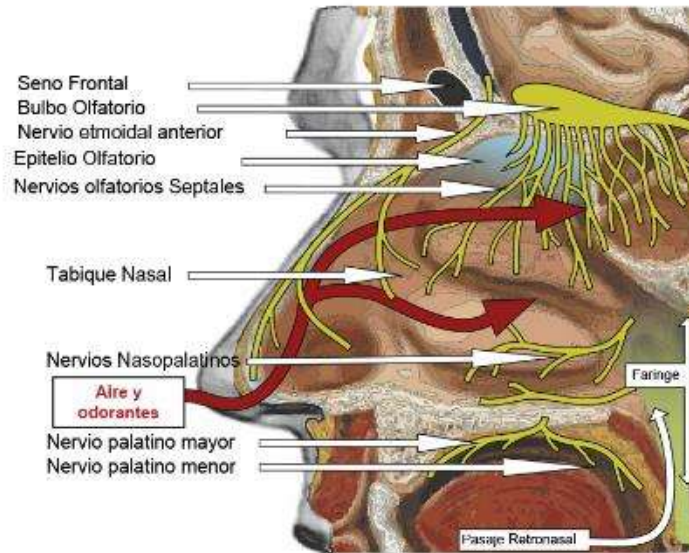


Figura 11. Aroma y receptores del gusto (96,99)

3.16. Análisis descriptivo sensorial

El análisis descriptivo es el método sensorial por el cual los atributos de un alimento o producto son identificados y cuantificados utilizando sujetos humanos que han sido entrenados específicamente para este propósito. El análisis puede incluir todos los parámetros del producto, o puede estar limitado a ciertos aspectos, por ejemplo, aroma, sabor, textura y retrogusto. En muchos casos, este método sensible de análisis descriptivo proporciona información que no se puede obtener por otros medios analíticos. Por ejemplo, el análisis del contenido de sal o el pH nos indica qué tan salado o agrio puede ser el sabor del producto. Tampoco es generalmente posible monitorear los cambios sutiles en la vida útil o la estabilidad de un empaque utilizando instrumentos analíticos. La única forma efectiva para monitorear los cambios complejos en la oxidación, rancidez o intensidad del sabor, así como la introducción de nuevos atributos que tan a menudo ocurren con el almacenamiento,

es mediante métodos de análisis descriptivo. Establecer un panel capacitado para realizar un análisis descriptivo no es un asunto trivial. Este método requiere que el panel sea entrenado y mantenido cuidadosamente bajo la supervisión de un profesional sensorial que tiene entrenamiento y experiencia en el método analítico que se aplica. Debido al costo (en dinero real y tiempo del personal) de la capacitación y el mantenimiento de un panel, así como la posible necesidad de inversión de capital de una instalación especial, la administración de la compañía, laboratorio o instituto deben proveer de un compromiso a largo plazo. Sin tal apoyo es casi imposible desarrollar y mantener con éxito el panel. Sin embargo, los beneficios de tener este importante método analítico generalmente superan las desventajas. Por esta razón, muchas empresas han encontrado que el método de análisis descriptivo es una parte esencial en su programa de evaluación sensorial. El método de análisis descriptivo cuantitativo (QDA) se basa en el principio de la capacidad de un panelista para verbalizar las percepciones de un producto de una manera confiable. El método incorpora un procedimiento formal de selección y entrenamiento, desarrollo y uso de un lenguaje sensorial, y la puntuación de los productos en ensayos repetidos para obtener una descripción cuantitativa completa (33).

4. JUSTIFICACIÓN DEL ESTUDIO

El Perú es considerado como uno de los 17 países mega diversos del mundo (100). Debido a ello es el centro de origen de varios cultivos importantes tales como la papa, maíz, cacao y ají.

Las especies del género *Capsicum* del Perú ofrecen una amplia gama de atributos sensoriales. Esto se explica por su gran diversidad debido a sus diferentes centros de origen, formas, tamaños, colores y una amplia gama de características tales como aromas, sabores y pungencia (101). Los capsaicinoides, responsables de la pungencia, son exclusivos en el reino Plantae, restringidos al género *Capsicum* (102) y son muy valorados por los consumidores. Aproximadamente el 90% de la pungencia se atribuye a la capsaicina y la dihidrocapsaicina (38). La gran diversidad de los ajíes nativos peruanos también representa un gran potencial para ingresar a mercados nacionales e internacionales. Esto se debe a un cada vez más creciente interés por nuevos sabores, alimentos e ingredientes diferenciados de alto valor nutricional y de beneficio para la salud. Los ajíes nativos peruanos, y productos derivados de estos, tanto frescos como procesados, responden a los cambios en las preferencias de los consumidores y al *boom* gastronómico que experimenta actualmente el Perú. Los ajíes nativos son un nicho de este mercado emergente.

En la actualidad, hay ejemplos de éxito en el comercio de ajíes nativos peruanos. El Perú es el octavo exportador y ocupa la posición número 27 de producción de

Capsicum a nivel mundial. Son de importancia económica las variedades nativas tales como el ají amarillo, ají panca, rocoto y ají escabeche, entre otros. En el año 2018 la exportación de *Capsicum* cerró con ganancias de US\$ 246 millones **(103)**.

En el año 2018 la UNALM y el INIA comenzaron a trabajar juntos en el marco del proyecto: «Consolidación de las colecciones INIA y UNALM de *Capsicum* nativos del Peru en una base de datos morfológica, genética, fisicoquímica y culinaria integrada y análisis de las prioridades de conservación y desarrollo» para consolidar ambas colecciones de ajíes nativos peruanos a partir de la información morfológica, genética y físico-química disponibles. Dicho proyecto no tiene contemplado realizar estudios metabolómicos de ajíes. Hay que destacar que la metabolómica es una alternativa muy importante porque se basa en el perfil de metabolitos secundarios que están íntimamente ligados con el fenotipo. Los resultados que se obtengan en el presente proyecto contribuirán a tener una caracterización más detallada (genómica y metabolómica) de 5 accesiones de ajíes nativos, lo cual permitirá una mejor discriminación entre especies. Tres de las accesiones pertenecen al complejo taxonómico de *Capsicum annum* (charapita, miscucho y dulce rojo) y dos al complejo taxonómico de *Capsicum baccatum* (tomatito rojo y ayuclo).

Otro aspecto importante de los frutos de ajíes son sus sabores y aromas, muy apreciados por los consumidores. Actualmente la evaluación de su calidad organoléptica es llevada a cabo por análisis descriptivo cuantitativo (QDA). Esta

técnica analítica está sujeta a errores debido a las propias subjetividades de los panelistas y su moderada reproducibilidad. Al reconocer las debilidades y problemas del QDA, es necesario buscar otros métodos alternativos o complementarios.

Para lograr este propósito sería de utilidad determinar metabolitos responsables del sabor y aroma. Hay que destacar que esta información no está disponible para los ajíes nativos y/o cultivados. El estado del arte de estas identificaciones reporta hasta la fecha un único estudio realizado por Eggink *et al.* donde se correlacionaron los atributos sensoriales del sabor con metabolitos volátiles y no volátiles. El modelo logró determinar los descriptores químicos del sabor, lo cual constituyó un punto de partida para una selección objetiva de sabores y aromas de una selección de pimientos (*C. annuum*) (37).

Las 5 accesiones de ajíes de la colección del INIA contienen atributos agromorfológicas, fisicoquímicos y sensoriales ya estudiados, dichas accesiones pasaron por un proceso previo de selección en los que participaron chefs y productores de ajíes (Empresa Agro Export Topara), quienes identificaron a estas accesiones como las más interesantes.

5. OBJETIVOS

5.1.Objetivo principal

Determinar el perfil metabolómico de las accesiones Ayucillo (*Capsicum baccatum*), Miscucho (*Capsicum chinense*), Tomatito rojo (*Capsicum baccatum*), Charapita amarillo (*Capsicum frutescens*) y Dulce rojo (*Capsicum chinense*) de la colección de ajíes nativos del INIA y evaluar la correlación de los metabolitos con los atributos sensoriales.

5.2.Objetivos específicos

- a. Obtener el perfil de metabolitos no volátiles (sabor) de las 5 accesiones de ajíes nativos por UHPLC-HRMS Orbitrap a partir de los frutos obtenidos de un cultivo de la localidad de Chincha
- b. Obtener el perfil de metabolitos volátiles (aroma) de las 5 accesiones de ajíes nativos por SPME-GC-MS a partir de los frutos obtenidos de un cultivo en la localidad de Chincha
- c. Generar un modelo de huella metabólica para la discriminación de las 5 accesiones a partir de los perfiles de metabolitos no volátiles de frutos obtenidos de un cultivo en la localidad de Chincha

- d. Validar el modelo de huella metabólica de la localidad de Chíncha en frutos de las 5 accesiones obtenidas de plantas crecidas en las localidades de Chiclayo y Huaral.
- e. Identificar los metabolitos volátiles y no volátiles que permitan predecir los atributos sensoriales de las 5 accesiones estudiadas

6. PROCEDIMIENTO EXPERIMENTAL

6.1. Diseño experimental

De acuerdo a este estudio se han identificado las variables independientes (variables manipuladas por el experimentador) y las variables dependientes (la medida de la respuesta). Asimismo, las variables “intervinientes” que modifican indirectamente la respuesta de las mediciones. Estas variables se mantendrán constantes para los 5 ajíes investigados. En la figura 12 se muestra el esquema del diseño experimental y la identificación de las variables.

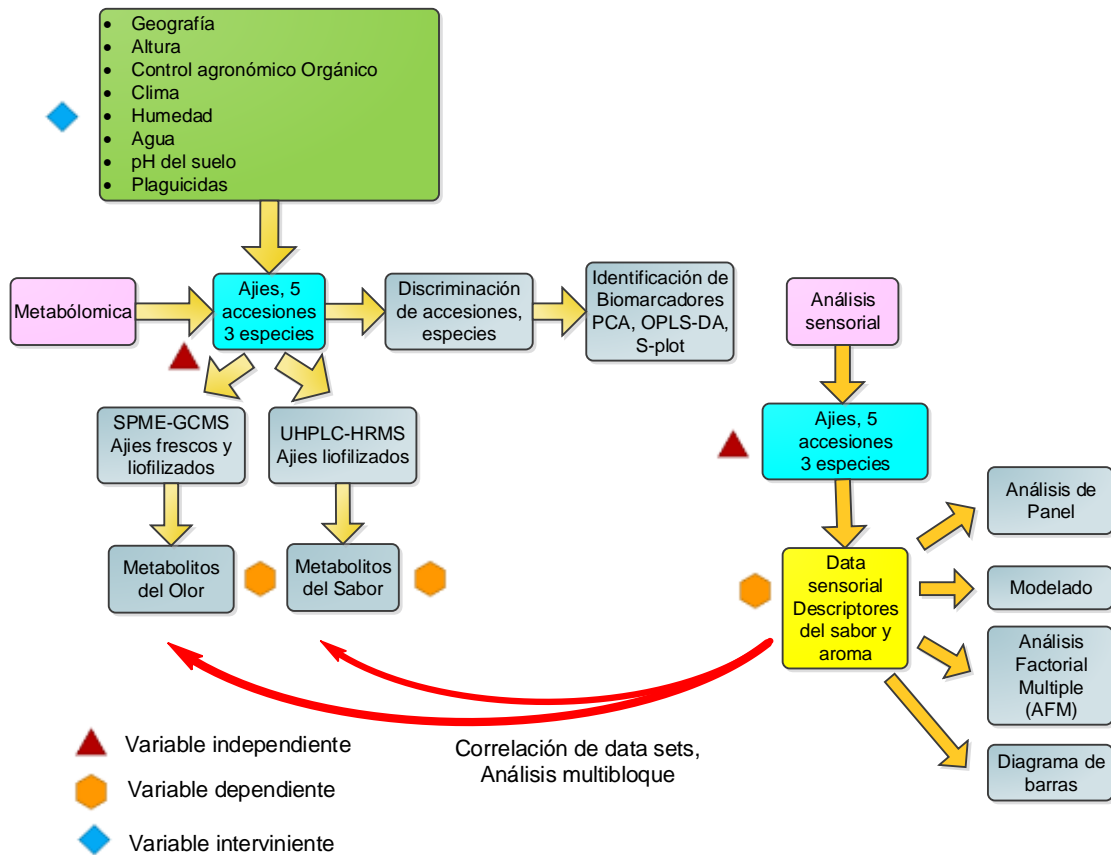


Figura 12. Esquema de diseño experimental e identificación de las variables

6.2.Población

De acuerdo al diseño experimental (Figura 12) se tienen tres poblaciones de ajíes correspondiente a 3 zonas de cultivos de la costa peruana (localidades), una población de ajíes de la localidad de Chinchá, con manejo orgánico del cultivo de *Capsicum* bajo las condiciones de la Empresa AgroExport Topará – Chinchá, ICA (Figura 13) y el manejo convencional del cultivo de *Capsicum* bajo las condiciones de las estaciones experimentales de Donoso – Huaral y Vista Florida – Chiclayo del INIA.



Figura 13. Producción de cultivares de ajíes orgánicos de la Empresa Topara bajo almacigado y vivero sembrío en bolsa (a y b), Ajíes charapita (c), miscucho (d), tomatito rojo (e) y dulce rojo (f) de la localidad de Chiclayo (104).

En el diseño experimental de campo las unidades experimentales fueron divididos por cercos vivos para evitar la polinización de los materiales de *Capsicum* (Figura 14). La población total de es de 600 ajíes por localidad.

6.3.Muestra

Para el análisis en el laboratorio, las muestras se colectaron frescas y luego fueron almacenadas. Para cada accesión de ají se colectó un peso en el rango de 0.1- 0.5 Kg dependiendo del rendimiento de la cosecha de cada material donde fueron colectados.

6.4.Operacionalización de variables

Las variables identificadas son las que se observan en la figura 12. Se han definido las variables según su relación de causa y efecto (variables dependientes e independientes). En la tabla 3 se dan la descripción de las variables dependientes: la variable SPME-GC-MS (metabolitos Volátiles), UHPLC-HRMS (metabolitos No Volátiles) y la variable análisis sensorial con su respectiva operacionalización. Se destaca las dimensiones, que se refiere a las características subdivididas de las variables. Los resultados de las variables van a depender de estas subcategorías. La columna de indicadores, se refieren a las propiedades de las variables susceptibles a ser medidas. La Escala es el patrón convencional de medición y el índice se refiere a la expresión del indicador.

6.5.Procedimiento y técnicas

Las figuras 15, 16 y 17 muestran los flujos de trabajo que se desarrollaron para abordar el análisis metabolómico, la identificación de biomarcadores para la discriminación de accesiones y especies con las técnicas estadísticas PCA, OPLS-DA, S-plot, la validación de los biomarcadores y la identificación de los descriptores químicos del sabor y aroma.

6.6.Consideraciones éticas

El estudio no involucra seres humanos, ni animales. Antes de iniciar el estudio, el proyecto pasó por el Comité Institucional de Ética de la Universidad Peruana Cayetano Heredia y conto con su aprobación. El proyecto ha sido registrado con código SIDISI 100052.

Tabla 3. Operacionalización de variables del estudio metabolómico de ajíes nativos

Variable	Definición conceptual	Dimensiones	Indicadores	Escala	Índice
SPME-GC-MS (Volátiles)	La microextracción en fase sólida seguida por cromatografía de gases y espectrometría de masas (SPME-GC-MS) se ha utilizado ampliamente para identificar y cuantificar los compuestos volátiles en alimentos. La precisión del método SPME-GC-MS se ve afectada por el efecto de los compuestos en la matriz de la muestra, tales como azúcares, ácidos orgánicos, aminoácidos, compuestos fenólicos y proteínas.	<ul style="list-style-type: none"> -Condiciones de absorción utilizando Fibra SPME DVB/CAR/PDMS -Condiciones de absorción utilizando Fibra SPME CAR/PDMS -Adquisición de data utilizando un cromatógrafo GC-MS - Condiciones cromatográficas: Columna cromatográfica apolar DB5MS, temperatura, tiempo de corrida. 	Número de características o perfil de metabolitos volátiles	Cuantificación de las Áreas de los picos cromatográficos	10 ⁵ -10 ⁸
UHPLC-HRMS (No Volátiles)	La espectrometría de masas (LC-HRMS) acoplada a espectrometría de masa de alta resolución se encuentra entre las técnicas analíticas más efectivas para la caracterización estructural y el análisis de productos alimenticios. Las mediciones de masa precisas han ganado gran popularidad debido a su gran capacidad para proporcionar información más completa sobre la masa molecular exacta, la composición elemental y la estructura molecular detallada de un compuesto dado.	<ul style="list-style-type: none"> -Sistema de solvente de extracción de la muestra -Adquisición de data utilizando cromatógrafo UHPLC acoplado a un espectrómetro de masas de alta resolución Orbitrap (HRMS) -Condiciones cromatográficas: sistema de gradiente de solventes, columna cromatográfica apolar C18, tiempo de corrida. -Modo de ionización ESI (+, -) -Resolución de masa -Rango de masa m/z 	Número de características o perfil de metabolitos no volátiles	Cuantificación de las Áreas de los picos cromatográficos	10 ⁶ -10 ¹¹
Análisis sensorial	El análisis sensorial es el examen de las propiedades organolépticas (sabor y aroma) de un producto realizado con los sentidos humanos.	<ul style="list-style-type: none"> -10 descriptores del sabor y aroma -Calidad del material de <i>Capsicum</i> -Susceptibilidades de los panelistas 	scores	Cuantificación de las intensidades relativas del sabor y aroma	0-10

CALLE																		BLOQUE 1									
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4		T4	T4	y	y	T5	T5	T5	y	y
y	y	T1	T1	T1	y	y	T2	T2	T2	y	y	T3	T3	T3	y	y	T4	T4	T4	y	y	T5	T5	T5	y	y	
CALLE																		BLOQUE 2									
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
y	y	T4	T4	T4	y	y	T3	T3	T3	y	y	T1	T1	T1	y	y	T5		T5	T5	y	y	T2	T2	T2	y	y
CALLE																		BLOQUE 3									
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T2	T2	T2	y	y	T3		T3	T3	y	y	T1	T1	T1	y	y
CALLE																		BLOQUE 4									
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y
y	y	T3	T3	T3	y	y	T5	T5	T5	y	y	T4	T4	T4	y	y	T1		T1	T1	y	y	T2	T2	T2	y	y

Figura 14. Diseño experimental en campo. Distancia entre planta y planta = 0,70 m; distancia entre surco y surco = 0.90 m; calle = 1.5 m, Ti (tratamientos) = 5; bloques = 4 (30 plantas/unidad experimental); y = cerco vivo (maíz a chorro continuo).

6.7. Flujo de trabajo de identificación de Biomarcadores

En la figura 15 se observa el flujo de trabajo para la identificación de biomarcadores a partir de los perfiles cromatográficos de las 5 accesiones de ajíes nativos. Luego de la deconvolución y alineación por MS-DIAL 3.96, se realizó la identificación de los metabolitos in Silico utilizando la base de datos del diccionario de productos naturales, haciendo una búsqueda por género y especie. También se realizó una búsqueda con las bases de MSFINDER 3.26, la base de datos de MS-DIAL y GNPS utilizando el software MetGem. Después de la identificación se realizó la combinación de las tablas de datos cromatográficas ESI (+,-). Esto permitió correlacionar ambas tablas de datos y filtrar los compuestos aductos.

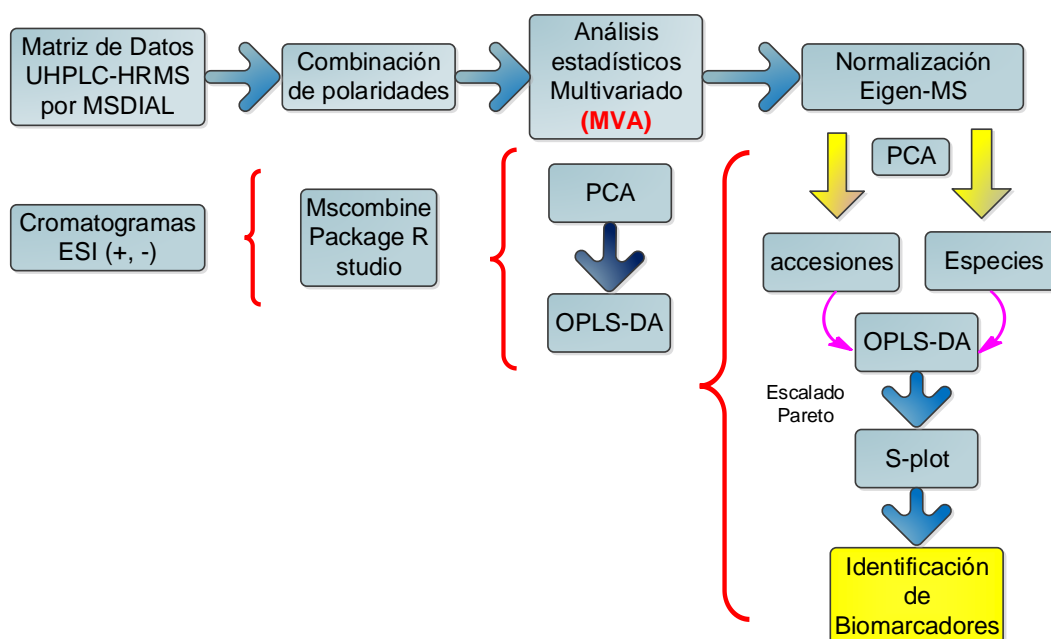


Figura 15. Flujo de trabajo para la identificación de biomarcadores por OPLS-DA y S-plot

La data cromatográfica fue sometida a un análisis estadístico multivariado tal como PCA y OPLS-DA. El PCA permitió visualizar la data cruda tal cual y formular

hipótesis, mientras que OPLS-DA permitió observar si las clases asignadas son discriminantes y juzgar la robustez del modelo. Luego, la normalización por el método EigenMS utilizando el software online NOREVA permitió eliminar la variación intraclase. La normalización se realizó por asignación de clases por accesiones y especies.

6.8. Flujo de trabajo de validación de Biomarcadores

En la figura 16 se observa el flujo de trabajo para la validación de los biomarcadores por predicción probabilística. Las muestras de ajíes provienen de tres cosechas en diferentes localidades: de la localidad 1 (Chincha), localidad 2 (Chiclayo) y localidad 3 (Huaral).

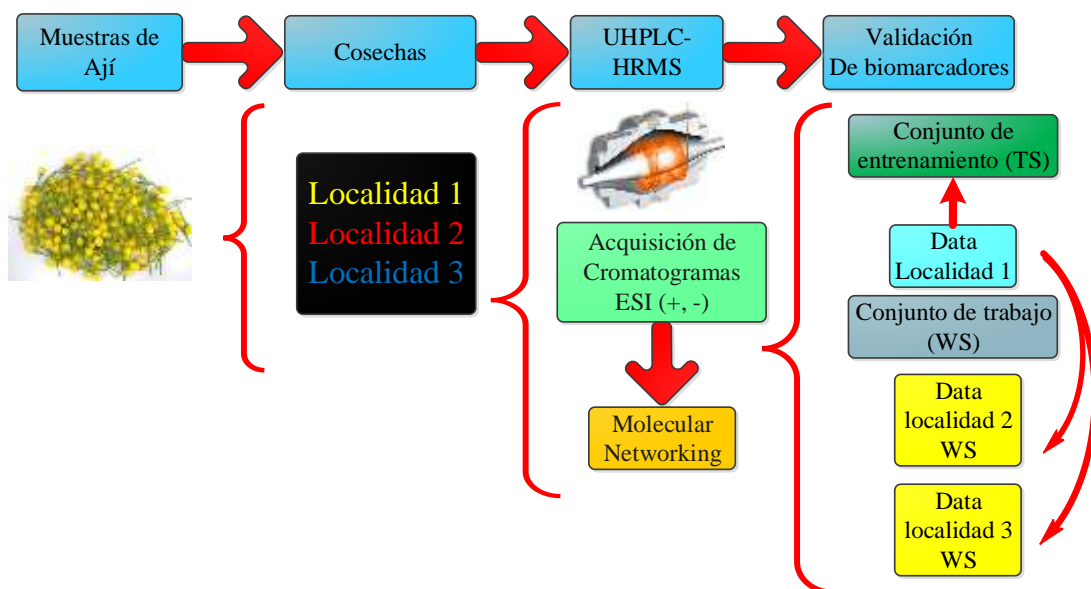


Figura 16. Flujo de trabajo para la validación de biomarcadores por predicción probabilística

Extractos etanólicos (80%) de muestras de ají liofilizados fueron analizados por UHPLC-HRMS. La data cromatográfica obtenida del análisis fue combinada y normalizada por el método EigenMS siguiendo el flujo de trabajo de la figura 15. De esta data cromatográfica se originaron tres tablas de datos cromatográficos de la localidad de Chincha, Chiclayo y Huaral. La data cromatográfica de la localidad 1 sirvió como conjunto de entrenamiento para desarrollar un modelo OPLS-DA, la data cromatográfica de la localidad 2 y 3 sirvieron como conjunto de trabajo o test. La validación se realizó utilizando el software SIMCA 15.

6.9. Flujo de trabajo para el análisis multibloque

En la figura 17 se muestra el flujo de trabajo para el análisis multibloque. El análisis multibloque es un método de reducción de la dimensionalidad de datos multivariado para la integración de múltiples tablas de datos ómicas (Package de R MixOmics). Tiene como objetivo maximizar la información común o correlacionada entre múltiples conjuntos de tablas de datos mientras que identifica de manera exitosa las variables ómicas clave (68). Entre los resultados del análisis se encuentran: el criterio de correlación interbloque que proporciona el valor de ajuste de R^2 de correlación de Pearson, gráfico de individuos, CIM o Hetmap y Circosplot. Del Circosplot o círculo de correlaciones se pueden exportar los resultados en una tabla de Excel con formato *.CSV para su análisis.

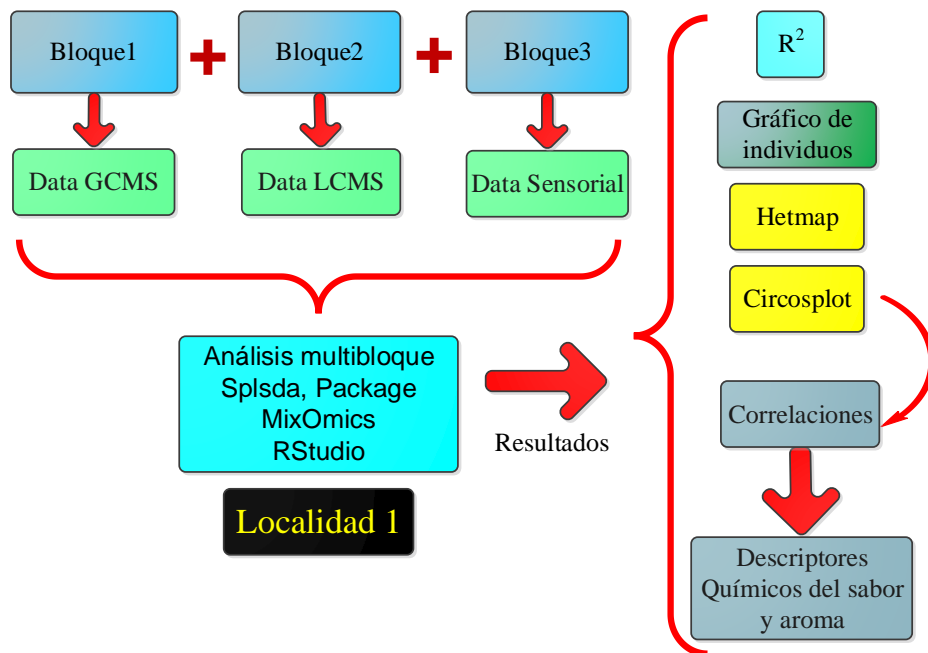


Figura 17. Flujo de trabajo para el análisis multibloque

6.10. Materiales y Métodos

6.10.1. Material de la Planta

Se trabajó con 5 accesiones de ajíes provenientes de la colección de accesiones de ajíes nativos del INIA (Figura 20). Las accesiones fueron sembradas en tres localidades en octubre del año 2016 y cosechadas entre enero y marzo del año 2017. Las tres localidades son las siguientes: 1) instalaciones de la Empresa Agroexport Topara ubicada en Chincha, 2) en la instalación experimental del INIA ubicado en Vista Florida – Chiclayo y 3) en la instalación experimental del INIA en Donoso – Huaral. La producción de los ajíes de Chincha tuvo un manejo agrícola orgánico mientras que para las otras 2 localidades tuvo un manejo agrícola convencional. Los ajíes colectados fueron lavados con agua corriente y secados con una toalla de algodón y luego guardados en refrigeración a -19 °C, para su posterior análisis sensorial efectuado por los panelistas y para su liofilización.

6.10.2. Liofilización de las muestras

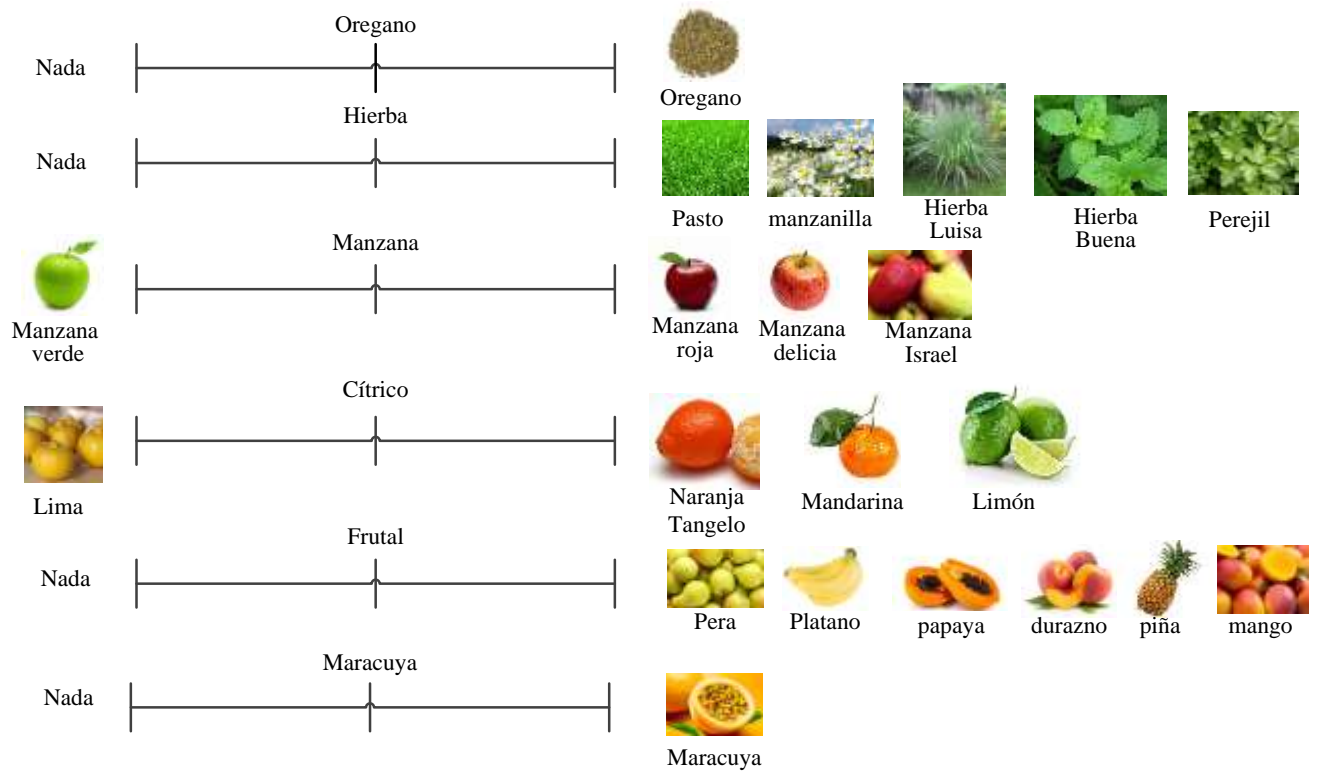
Las muestras frescas fueron secadas por liofilización a una presión de 0.5 mbar y temperatura de -40 °C por espacio de 2 días. Se liofilizaron aproximadamente 0.5 kg de material fresco (fruto entero). El rendimiento de liofilización fue aproximadamente del 10%. Luego fueron molidas hasta obtener un polvo fino (Figura 18).



Figura 18. Ajés enteros liofilizados y molidos

6.10.3. Análisis sensorial

El análisis sensorial fue llevado a cabo utilizando 135 ajés por localidad (405 ajés) en el que participaron 9 catadores entrenados. Para el experimento se utilizó el QDA (análisis descriptivo cuantitativo) como metodología de obtención de los puntajes sensoriales o score. Para realizar el análisis se utilizaron diez descriptores del sabor y aroma (4 descriptores de gusto y 6 descriptores de aroma) (Figuras 19 y 21). Estos fueron establecidos previamente durante el entrenamiento. Cada catador evaluó cada accesión de ají por triplicado. Para cada evaluación se entregó la muestra codificada, implementos para realizar el corte, un alimento de sabor neutro y agua para eliminar del paladar los sabores residuales luego de cada evaluación.



Mínimo

Máximo

Figura 19. Descriptores del olor (orégano, hierba, manzana, cítrico, frutal, maracuyá), se muestran las anclas mínimo y máximo con sus respectivos atributos.

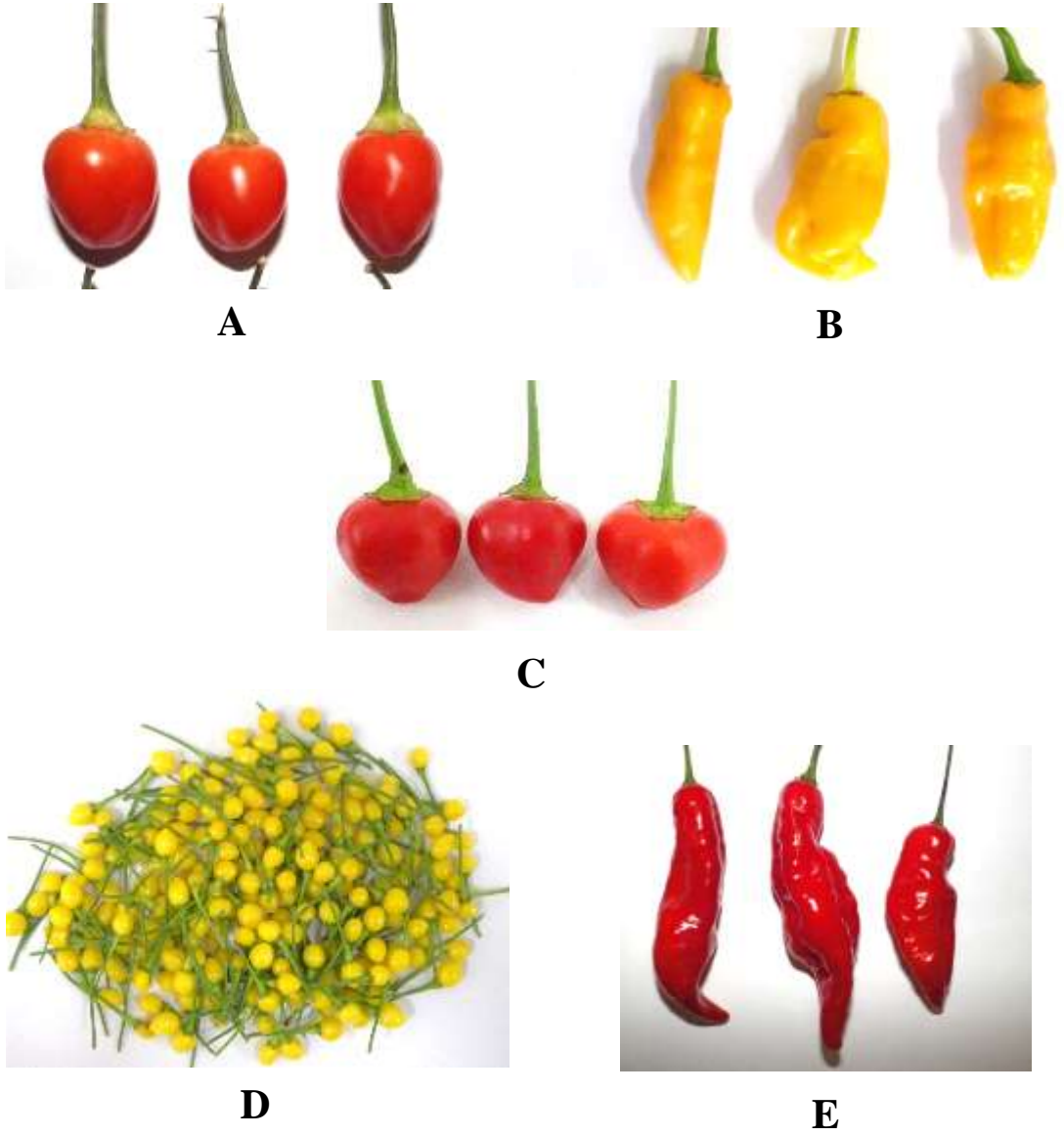


Figura 20. Acciones de Ajíes Nativos: A) Ayuello (*Capsicum baccatum*), B) Miscucho (*C. chinense*), C) Tomatito Rojo (*C. baccatum*), D) Charapita (*C. frutescens*), E) Dulce Rojo (*C. chinense*).

La evaluación se llevó a cabo en cabinas individuales. Dado que el color de la muestra es una propiedad sensorial que puede influenciar la evaluación y la percepción subjetiva de los atributos, se enmascaro utilizando 3 luces de color rojo, violeta o verde, dependiendo del color de cada muestra de ají (44). Al recibir la muestra, el panelista realizó un corte longitudinal para evaluar los descriptores del aroma con el sentido del olfato. Luego de cortar y colocar en el paladar un pedazo de la parte media del pericarpio se evaluaron los descriptores del sabor. El panelista indicó su calificación en una escala no estructurada de 10 cm que consistió de una línea trazada entre dos puntos extremos (máximo y mínimo de intensidad del atributo) y un punto central. Los valores obtenidos para cada atributo fueron promediados y luego fueron utilizados para la elaboración del perfil sensorial de cada ají.

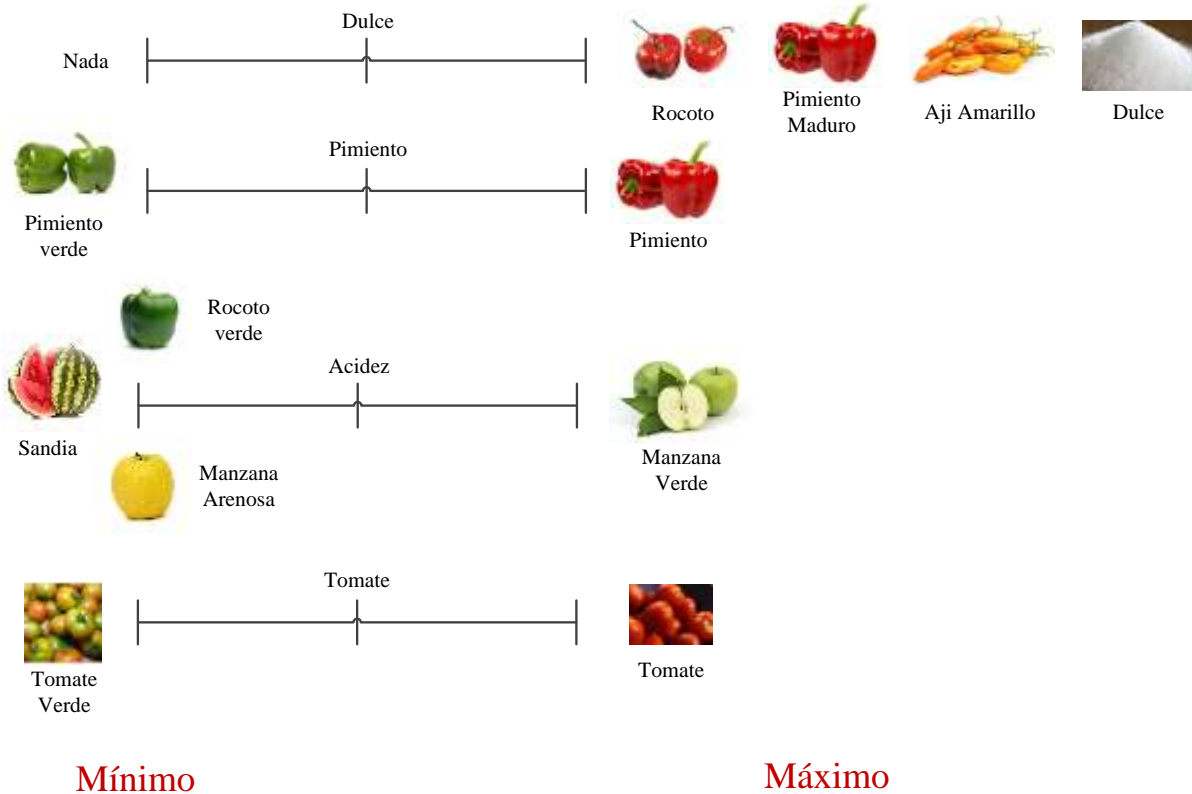


Figura 21. Descriptores del gusto (Dulce, Pimiento, acidez, Tomate), se muestran las anclas mínimo y máximo con sus respectivos atributos.

6.10.4. Análisis proximal

Se realizó el análisis proximal de las muestras de ají liofilizado de la localidad de Chincha. En la tabla 15 se dan los resultados de las determinaciones proximales e instrumental. A continuación, se describen las metodologías de las determinaciones:

6.10.5. Determinación de Humedad

En una cápsula se pesó 5 g de ají fresco (pericarpio, placenta y semillas). La muestra fue sometida a desecación a 105 °C en una estufa electrónica automática por 4

horas. Luego, se retiró de la estufa y se dejó enfriar en un desecador por 30 min hasta alcanzar un peso constante. Transcurrido ese tiempo, se pesó la cápsula en una balanza analítica (Ohaus, Adventurer®). La cantidad de humedad fue determinada por la pérdida de peso de la muestra. El procedimiento se efectuó por triplicado y el porcentaje de humedad fue calculado usando la siguiente fórmula:

$$\% \text{ humedad} = \frac{(\text{Peso}_{\text{Muestra fresca}} - \text{Peso}_{\text{Muestra seca}})}{\text{Peso}_{\text{Muestra fresca}}} \times 100$$

6.10.6. Determinación de grasas

Se pesó 5 g de muestra de fruto seco liofilizado (pericarpio, placenta y semillas) los que luego fueron colocados en cartuchos de papel filtro. Luego se colocaron en la cámara de extracción del equipo extractor Soxhlet y se agregó éter de petróleo como solvente de extracción. El sistema fue calentado hasta una temperatura de 80 °C durante 4 horas (Figura 22). Transcurrido el tiempo de extracción, el solvente fue evaporado a presión reducida a una temperatura de 40 °C utilizando un rotavapor (Büchi, modelo Rotavapor II). Luego de evaporar el solvente, se colocó cada balón en una estufa electrónica a 40 °C por 30 minutos, después de retirar los balones de la estufa, se colocaron en un desecador y finalmente se pesó el balón conteniendo la grasa de ají. Se determinó la cantidad de grasa presente en la muestra por diferencia de pesos respecto al balón vacío. Para el cálculo del porcentaje de grasas, se utilizó la siguiente fórmula:

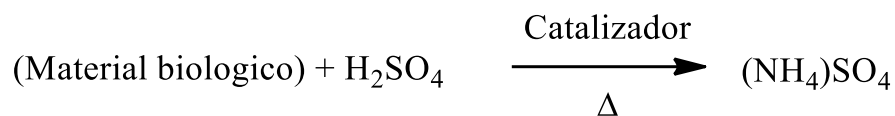
$$\% \text{ grasas}_{\text{base seca}} = \frac{(\text{Peso}_{\text{balón con grasa}} - \text{Peso}_{\text{balón}})}{\text{Peso}_{\text{Muestra seca}}} \times 100$$



Figura 22. Extracción de grasas de ají liofilizado utilizando el equipo extractor Soxhlet

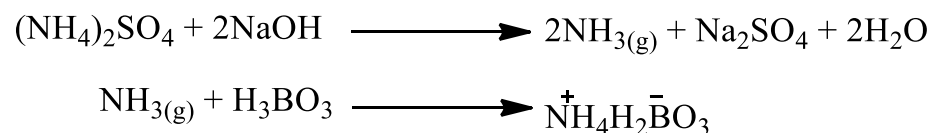
6.10.7. Determinación de Proteínas

La proteína cruda fue determinada por el método de Kjeldahl, convirtiendo todo el Nitrógeno orgánico de 1.00 g de muestra seca y molida en Nitrógeno amoniacal (NH_3). En la digestión Ácida La muestra seca (1.0 g) fue sometida a digestión, introduciéndola con cuidado en un tubo de digestión de vidrio (VELP® Scientifica) que contenía previamente una mezcla homogenizada de 4.0 g sulfato de sodio (Na_2SO_4) y 0.8 g sulfato de cobre ($\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$), el procedimiento fue realizado con cuidado de tal manera que no quede muestra pegada en las paredes del tubo, agregándole finalmente 14 mL ácido sulfúrico concentrado P.A. (98%). El procedimiento fue realizado por triplicado y se agregó un blanco que constó de todos los reactivos y en lugar de la muestra se colocaron pedazos de cerámica. Los balones fueron colocados en un sistema de hornillas de digestión semiautomática, y se fue subiendo la temperatura lentamente hasta llegar a más

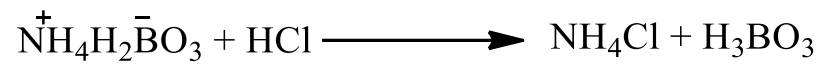


de 250 °C, se dejó con esta temperatura constante hasta que la mezcla inicialmente color celeste reaccionó por completo liberando vapores, para que finalmente tome una coloración verde esmeralda, luego de aproximadamente 2 a 3 horas.

Finalizada la digestión, se dejó enfriar el tubo de digestión conteniendo el producto y se dispuso en la unidad de destilación automática VELP® SCIENTIFICA UDK 129, alimentando al equipo con agua destilada tipo I y una solución de NaOH al 50% para la neutralización de la mezcla, el equipo se programó para destilar por 5 minutos, con un volumen igual a 40 mL de NaOH, además para recibir el destilado se colocó un matraz Erlenmeyer de 250 mL con una solución de ácido bórico (H₂BO₃) al 4% al que se le agregó 3 gotas del indicador verde de bromocresol, recuperándose unos 150 mL de solución de color azul en cada matraz. La reacción de neutralización fue la siguiente:



Los destilados obtenidos fueron valorados con una solución de HCl, previamente estandarizada 0.10472 N, (con el estándar primario tris-hidroximetil aminometano, PM = 121.14 g), hasta la variación del color inicial azul hasta amarillo. Los volúmenes de gasto obtenidos de HCl fueron anotados.



El porcentaje de proteína fue calculado utilizando las siguientes ecuaciones:

$$\% \text{Nitrógeno Kjeldahl} = \frac{[(V_s - V_B) \times N \times 14.01]}{10W}$$

$$\% \text{Proteína Total} = \% \text{Nitrógeno Kjeldahl} \times F$$

Donde:

V_s : Volumen gastado en la titulación del destilado de la muestra

V_B : Volumen gastado en la titulación del destilado del blanco

N : Normalidad del HCl usado en la titulación

W : Peso de la muestra

F : Factor de conversión de 6.25

6.10.8. Determinaciones Instrumentales

6.10.8.1. Determinación del pH

Se determinó el pH de los extractos acuosos de ajíes liofilizados. Para preparar los extractos se pesaron en un tubo de centrifuga, 1g de ají liofilizado y se agregó 30 mL de agua tipo I (18 MΩ.cm). El extracto se llevó a agitación en un vortex por 1 min. Posteriormente se centrifugó a 5000rpm por 15 min y temperatura de 4°C. Se midió el pH de las soluciones de sobrenadante utilizando un pH metro electrónico previamente calibrado con soluciones buffers a pH 4, 7 y 10. Las mediciones fueron realizadas por triplicado.

6.10.8.2. Determinación de los grados Brix

Los grados brix se determinaron utilizando soluciones de extracto acuoso de ají liofilizado. Para la preparación de los extractos se pesó 1g de ají liofilizado y se agregó 30 mL de agua tipo I. El extracto obtenido se filtró con papel de filtro. Se utilizó el filtrado para medir los grados Brix de los ajíes. Los grados Brix se midieron utilizando un refractómetro PAL-1 ATAGO de rango (0-53%). Las determinaciones fueron realizadas por triplicado.

6.10.9. Análisis estadístico

La data sensorial cruda fue evaluada con las pruebas de distribución normal de Shapiro Wilk, Anderson-Darling, lilliefors y Jarque-Bera. También, un análisis de Panel permitió evaluar la capacidad de los panelistas para distinguir entre los descriptores del sabor y aroma. Ambos análisis se realizaron utilizando el software XLSTAT de Excel. Con la información de la media y desviación estándar de la data sensorial cruda, utilizando el software (Autograph v.4.0.12, 30-day trial), se construyó un modelo de puntajes sensoriales (scores), para generar una población de $N = 300$ datos con distribución normal para cada descriptor sensorial, con lo cual se diseñó un modelo sensorial. La data sensorial optimizada fue evaluada por PCA, OPLS-DA, test de permutaciones, área bajo la curva de los gráficos ROC y CV-ANOVA para su validación interna utilizando SIMCA 15. La confiabilidad del modelo (figura 45) se realizó utilizando la data sensorial cruda. También se realizó el Análisis Factorial Múltiple (AFM) de la data sensorial optimizada. El AFM se realizó utilizando los Package FactorMineR y Factoextra del software de Rstudio. El análisis ANOVA y análisis Post Hoc de prueba de Tukey de la data sensorial optimizada permitió observar las diferencias significativas entre las intensidades relativas de los descriptores por cada accesión. El ANOVA se realizó utilizando el software XLSTAT de Excel.

6.10.10. Análisis por UHPLC-HRMS Orbitrap

Se trabajó con un cromatógrafo líquido UHPLC Dionex Ultimate 3000 Thermo Scientific equipado con un detector de arreglo de diodos, detector UV y detector de

aerosol de carga (CAD). Una columna Waters Acquity UHPLC BEH C18 Particle size 1.7 μm , de dimensiones 2.1x100 mm equipado con un guarda columna ULTRA Holder de la marca phenomenex para columna UHPLC 2.1 a 4.6 mm ID. El cromatógrafo líquido acoplado a un espectrómetro de masas equipado con un analizador de masas de trampa de iones Orbitrap LTQ XL- ETD Thermo Scientific y sistema de ionización por electrospray ESI en los modos positivo y negativo respectivamente. La separación cromatográfica fue llevada a cabo utilizando las siguientes condiciones cromatográficas: sistema A: H_2O + 0.1% FA; sistema B: ACN + 0.1% FA; de 0→5 min 95% A, 5% B; 5→12.2 min, 5% A, 95% B.

6.10.11. Perfilado UHPLC-HRMS Orbitrap

En microviales de centrifuga eppendorf de 1.7 mL, se pesaron 10 mg de ají liofilizado por triplicado. La extracción de los metabolitos fue llevada a cabo utilizando metanol HPLC al 80% en agua de tipo I (18 $\text{M}\Omega\cdot\text{cm}$), según el siguiente procedimiento:

- 1) A 10 mg de muestra pesada se añadió 800 μL de metanol 80%, pasando luego 1 minuto por vortex y 10 minutos por ultrasonido y pasando nuevamente 1 min por vortex.
- 2) Se centrifugó a 10000 rpm por 10 minutos a 4 °C.
- 3) Luego se tomó 600 μL del sobrenadante y se pasó a otro vial eppendorf.
- 4) Se repitió los pasos 1 a 3 para la segunda extracción, obteniéndose un volumen total de 1200 μL de extracto.
- 5) De cada batch de muestra por triplicado se preparó una muestra “MIX” que

consistió en tomar 200 μL de cada vial y hacer una mezcla en un vial nuevo obteniéndose un volumen de 600 μL .

También se preparó muestras para el blanco y el control de calidad (QC). El blanco consistió de una muestra de disolvente metanol al 80% acuoso y el QC se preparó tomando 100 μL de cada mix, es decir: $\text{Qc1}(100\mu\text{L}) + \text{Qc2}(100\mu\text{L}) + \dots\text{Qc15}(100\mu\text{L}) = 1500 \mu\text{L}$. La muestra QC y blanco se inyectaron de manera intermitente. La figura 23 muestra el orden de inyección de las muestras.

Orden de inyección	código de muestra	muestra
1	20180119-FE1-24-blk1	blk
2	X20180119.FE1.QC.Total1	QC
3	X20180119.FE.1.24.1	Charapita-c
4	X20180119.FE.1.24.2	Charapita-c
5	X20180119.FE.1.24.3	Charapita-c
6	X20180119.FE.1.24.4	Charapita-c
7	X20180119.FE.1.24.5	miscucho-c
8	X20180119.FE.1.24.6	miscucho-c
9	X20180119.FE.1.24.7	miscucho-c
10	X20180119.FE.1.24.8	miscucho-c
11	X20180119.FE.1.24.9	ayuclo-c
12	X20180119.FE.1.24.10	ayuclo-c
13	X20180119.FE.1.24.11	ayuclo-c
14	X20180119.FE.1.24.12	ayuclo-c
15	20180119-FE1-24-blk2	blk
16	X20180119.FE1.QC.Total2	QC
	.	.
	.	.
	.	.
	X20180119.FE1.25.33	tomatito -h
	X20180119.FE1.25.34	tomatito -h
	X20180119.FE1.25.35	tomatito -h
	X20180119.FE1.25.36	tomatito -h
71	20180119-FE1-24-blk6	blk
72	X20180119.FE1.QC.Total6	QC

Figura 23. Orden de inyección de las muestras, el blanco (Blk) y el control de calidad (QC) para perfilado por UHPLC-HRMS Orbitrap

6.10.12. Adquisición de la data

El software de manejo Xcalibur permitió adquirir y procesar la data automáticamente, así como el sistema de control del instrumento UHPLC-HRMS. El Programa MS Tune, nos permitió ajustar la adquisición de la data para el detector MS. Las muestras preparadas se cargaron en el automuestreador y en una ventana de configuración de secuencia se creó una secuencia en una fila y se adquirió un archivo de datos sin procesamiento de las muestras. Después de crear un método de procesamiento, se ejecutó la secuencia y adquirió un conjunto de archivos de datos sin procesar (tipo de archivo RAW) y luego el sistema de datos utilizó el método del instrumento, la posición y el volumen de inyección para adquirir los archivos de datos.

6.10.13. Procesamiento de Data UHPLC-HRMS

El tipo de data cromatográfica adquirida por UHPLC-HRMS fue del tipo centroide data dependent (Orbitrap LTQ XL- EDT), para procesar la data se utilizó una Laptop CORE i7-7700HQ 7th Gen, 37 MB de memoria, RAM. La data cromatográfica fue convertida del formato Thermo *.RAW a formato compatible *.abf fue con el software Abf Convert, y la deconvolución y alineación de los cromatogramas fue realizado con el software MS-DIAL(105). Para la identificación de los metabolitos se utilizó el software MSFINDER (106), utilizando su base datos interna KNApSAcK (Natural products), NANPDB (Natural products), PlantCyc (Plant),

UNPD (Natural products), ChEBI (Biomolecules) y LipidMAPS (Lipids), la base de datos MS2 pública (Publics DBs para ESI (+) y ESI (-)), la base de datos de Dictionary of Natural Products (DNP) 26.2 CRS Press y GNPS con la ayuda del software Metgem (107). La data fue filtrada utilizando los índices de % RSD>30%, t-test>0.05 y ratio Blk/QC>1.5 (que se eliminaron de la matriz de datos). También se procedió a la normalización de la data y escalado pareto. Para combinar la data UHPLC-MS de los modos de ionización ESI (+) y ESI (-), se utilizó el Package MScombine del software R y Rstudio (66). El análisis multivariado PCA, OPLS-DA y la identificación de biomarcadores por análisis del diagrama S-plot fueron realizados utilizando el software SIMCA-P 15.

6.10.14. Análisis por SPME-GC-MS

En la extracción de compuestos volátiles se trabajaron con dos fibras SPME de polímero marca Supelco (SPME Fiber Assembly 85 μ m CAR/PDMS, Stableflex 24GA, Manual Holder, 3pK (celeste)), y fibra SPME 50/30mm DVB/CAR/PDMS Stableflex 24GA, Manual Holder, 3Pk (ploma), para las muestras de ajíes frescos y liofilizados respectivamente. Antes de empezar con la absorción de volátiles las fibras fueron acondicionadas en el puerto de inyección GC-MS a 250 °C por 10 min. Para extraer los metabolitos volátiles de los ajíes frescos, se siguió el flujo de trabajo mostrado en la figura 24. Se tomaron entre 4 a 5 ajíes, los cuales se cortaron en pequeños trozos y luego se apisonaron en un mortero hasta hacer una pasta. De esta pasta se pesó aproximadamente 1.3 g en un vial SPME. Luego de sellar el vial

con su septa y tapa rosca, se introdujo la fibra SPME frente al espacio de cabeza y se dejó un tiempo de equilibrio de 10 min. A continuación, se llevó el sistema a un baño de agua con ultrasonido a 60 °C por un espacio de 30 min, e inmediatamente después, se retiró la fibra y se introdujo al sistema de inyección GC-MS para la desorción de los metabolitos volátiles. Por otro lado, para la extracción de metabolitos volátiles de las muestras de ajíes liofilizados se pesó 1.3g de ají seco en un vial SPME, Luego de sellar el vial con su septa y tapa rosca, se introdujo la fibra SPME ploma frente al espacio de cabeza y se dejó un tiempo de equilibrio de 60 min. Después, la fibra se introdujo al sistema de inyección GC-MS para la extracción de los metabolitos volátiles. El análisis GC-MS fue llevado a cabo utilizando un cromatógrafo de gases Agilent Technologies 7890A acoplado a un espectrómetro de masas 5975C. Los compuestos volátiles fueron separados utilizando una columna capilar apolar Agilent J&W DB-5ms 60mx0.25mmx0.25 μm , con el siguiente programa de temperatura: a 45 °C por 0 min, 3 °C/min a 100 °C por 0 min, 2 °C/min a 170 °C por 0 min, 10 °C/min a 230 °C por 0 min, 5 °C/min a 250 °C por 5 min. Post run 2 min a 300 °C, gas portador, helio (1 mL/min), volumen de inyección de 0.1 μL , modo splitless, run time: 68.333 min; rango de adquisición de masa: 45-300 m/z a 1s de intervalo (108,109).

6.10.15. Adquisición y procesamiento de la data SPME-GC-MS

La adquisición de la data GC-MS fue llevada a cabo con el software MassHunter de Agilent. Luego de la adquisición los archivos se convirtieron de formato *. D a

formato *. CDF. El procesamiento de la data, la deconvolución y alineación fue realizada utilizando el software MS-DIAL(105), y el análisis estadístico de PCA y OPLS-DA fue realizado utilizando el software SIMCA-P 15. La identificación de los metabolitos volátiles fue llevada a cabo utilizando el índice de Kovats. Los índices de Kovats de los metabolitos fueron calculados a partir de los tiempos de retención de los estándares de alcanos C8-C20. Para las identificaciones putativas se utilizaron las mayores librerías de base de datos NIST 14 (mainlib, replib), la base de datos flavor2 y la base de datos pública All-public GC-MS de MS-DIAL. Los compuestos identificados fueron reportados en base a su similaridad de Dot Product (110).

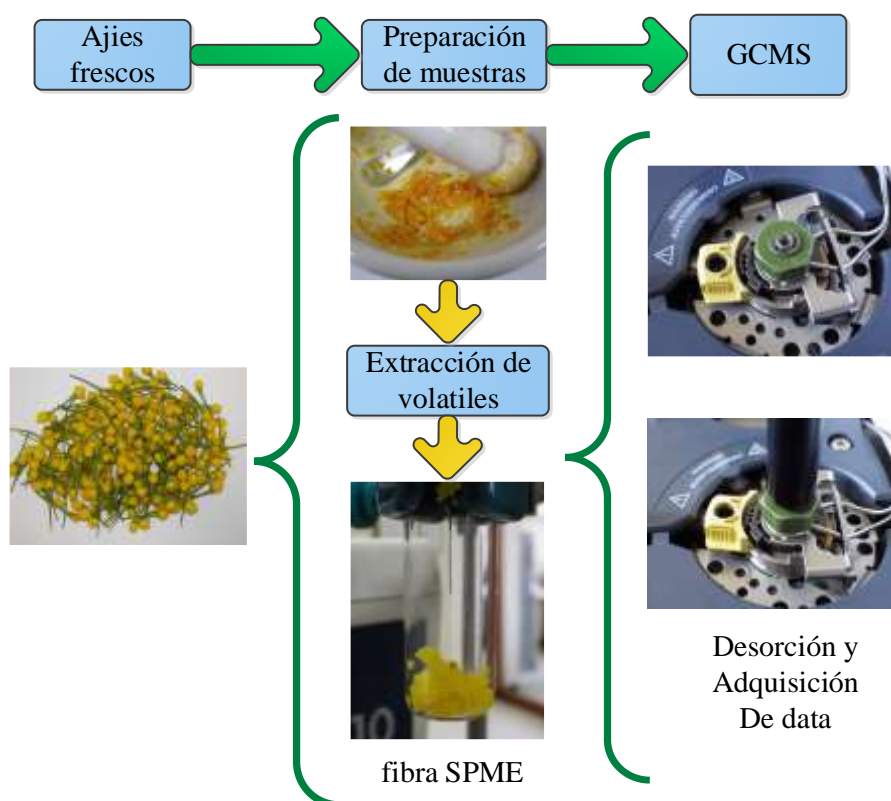


Figura 24. Flujo de trabajo para el análisis de volátiles por SPME-GC-MS

7.RESULTADOS

7.1.Perfil de metabolitos no volátiles de los ajíes de Chincha

En las figuras 25-29 se observan los cromatogramas LCMS de los extractos metanol-agua (MetOH-H₂O, 80%) de las 5 accesiones ajíes nativos en los modos de ionización ESI (+, -). Estos fueron obtenidos utilizando una columna apolar RP18. En el modo ESI (+) se observan tres características alrededor de los tiempos de retención de 0.69, 7.27 y 7.72 min, con relaciones masa/carga con valores de 146.08, 306.2072 y 308.2219, respectivamente, los cuales son los metabolitos de mayor área relativa para el sistema en cuestión. Por otro lado, en el modo ESI (-) se observan 2 características alrededor de los tiempos de retención 0.72 - 0.90 min con relación masa/carga de 191.0133, 191.0134 y 133.0099. Al inicio del cromatograma (0-2 min), son retenidos los metabolitos polares (ácidos orgánicos o azúcares), mientras que a RT entre 7.27 y 7.72 min, se han anotado como Capsaicina y Dihidrocapsaicina. El grueso de metabolitos secundarios son metabolitos semipolares y apolares que se encuentran como características minoritarias.

El tratamiento de la data fue realizado en ambos modos de ionización individualmente utilizando el software MS-DIAL. Las condiciones de trabajo por MS-DIAL en el modo de ionización ESI (-) fueron las siguientes: Precisión de masa: tolerancia MS1 = 0.01, tolerancia MS2 = 0.4, tiempo de retención inicial = 0.3 min, tiempo de retención final = 11.5 min, rango de masa inicial = 100 Da,

rango de masa final = 2000 Da. Amplitud = 335709.003. Mientras que para la data cromatográfica en el modo de ionización ESI (+) fueron las siguientes: Precisión de masa: tolerancia MS1 = 0.01, tolerancia MS2 = 0.4, tiempo de retención inicial = 0.3 min, tiempo de retención final = 11.5 min, rango de masa inicial = 100 Da, rango de masa final = 2000 Da. Amplitud = 498059.2125.

Después de la deconvolución se procedió con la identificación de las características o picos cromatográficos utilizando el software MS-FINDER (ver sección 13.3). Luego, se procedió con la combinación de la data ESI (+, -) utilizando el software estadístico de R y el Package MSCombine. En las tablas 57-66 se reportan las áreas relativas de las principales 40 características identificadas para las 5 accesiones de los ajíes de Chincha (ANEXO 1).

La tabla 4, muestra el número de características encontradas para las 5 accesiones de ajíes. En la segunda columna de la tabla 4 se indica el número total de características (deconvolucionados) y en la tercera columna se indican las características identificadas con bases de datos.

Tabla 4. Características totales e identificadas para las cinco accesiones de ajíes

Accesión	Características Totales	Características identificadas
Charapita	384	194
Miscucho	322	193
Tomatito	245	131
Ayucllo	256	130
Dulce Rojo	275	160

Las figuras 30-34 muestran los diagramas de barras para las seis principales características de las cinco accesiones. Para la accesión Ayucllo se observa la mayor área relativa en ácido cítrico. Las accesiones Charapita y Tomatito rojo presentan la mayor área relativa en Capsaicina. También las accesiones Charapita y Miscucho presentan las mayores áreas relativas en Dihidrocapsaicina.

Para determinar qué diferencias y similitudes existen entre los 5 metabolomas estudiados se procedió con la alineación y la deconvolución de los cromatogramas con el software MSDIAL. Los cromatogramas fueron obtenidos de 5 extractos metanólicos 80% de las muestras de ajíes, donde cada extracto tuvo 4 réplicas, además de trabajar con 6 muestras de control de calidad (QC) y 6 muestras del blanco (metanol 80%) para las tres localidades. Luego se procedió a realizar el análisis estadístico de PCA y posteriormente análisis discriminante OPLS-DA y gráficos S-plot con el objetivo de encontrar los metabolitos específicos ó huella metabólica para cada especie estudiada que permiten diferenciarlos. Esto se discute con más detalle en la sección 7.6, pag,154. En esta sección solo se muestran los resultados del análisis de los cromatogramas (análisis del perfilado) individualmente por cada muestra de ají nativo.

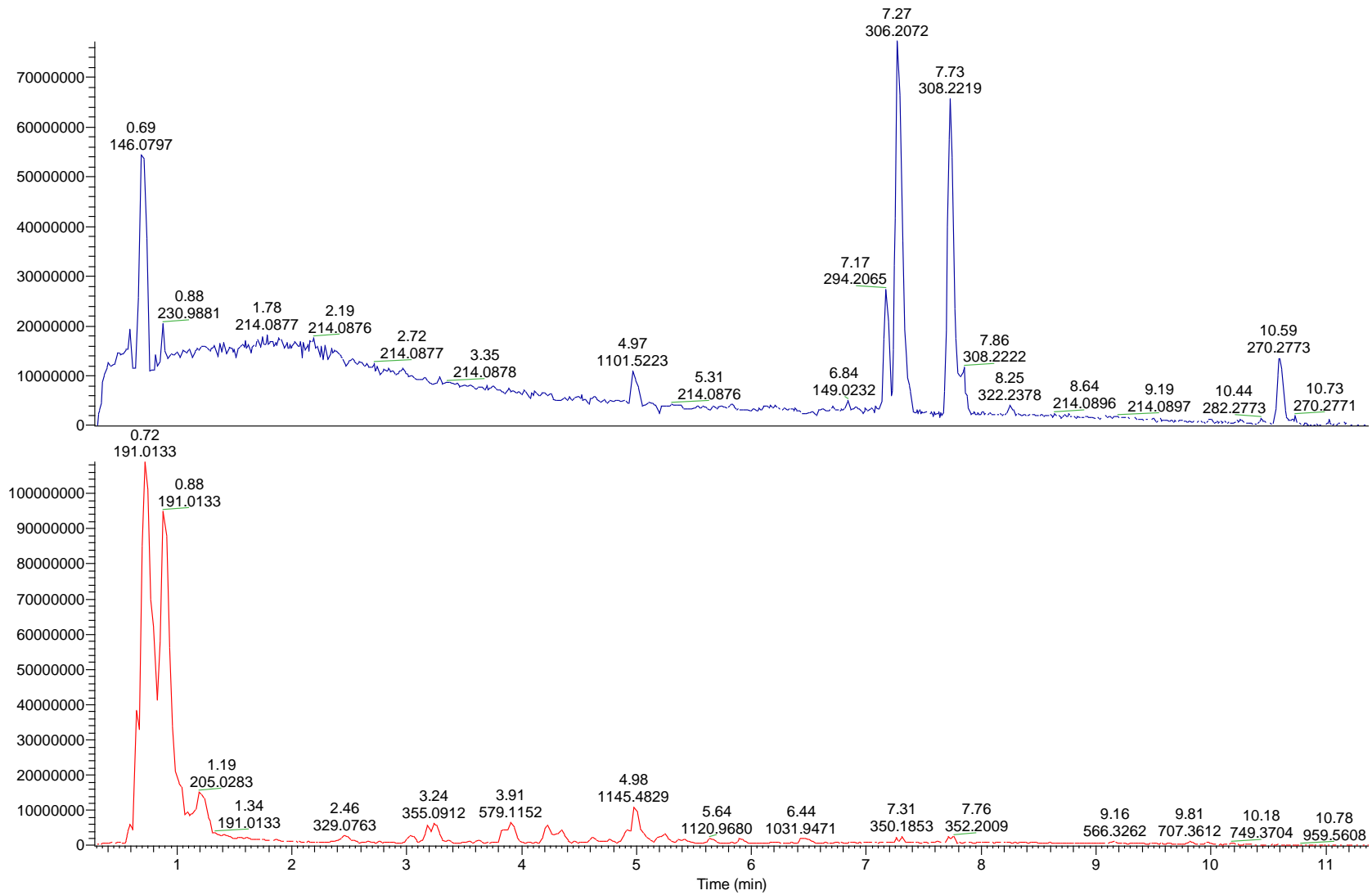


Figura 25. Perfil cromatográfico UHPLC-HRMS de los modos de ionización ESI (+, -) de la accesión Ayucllo (pico base)

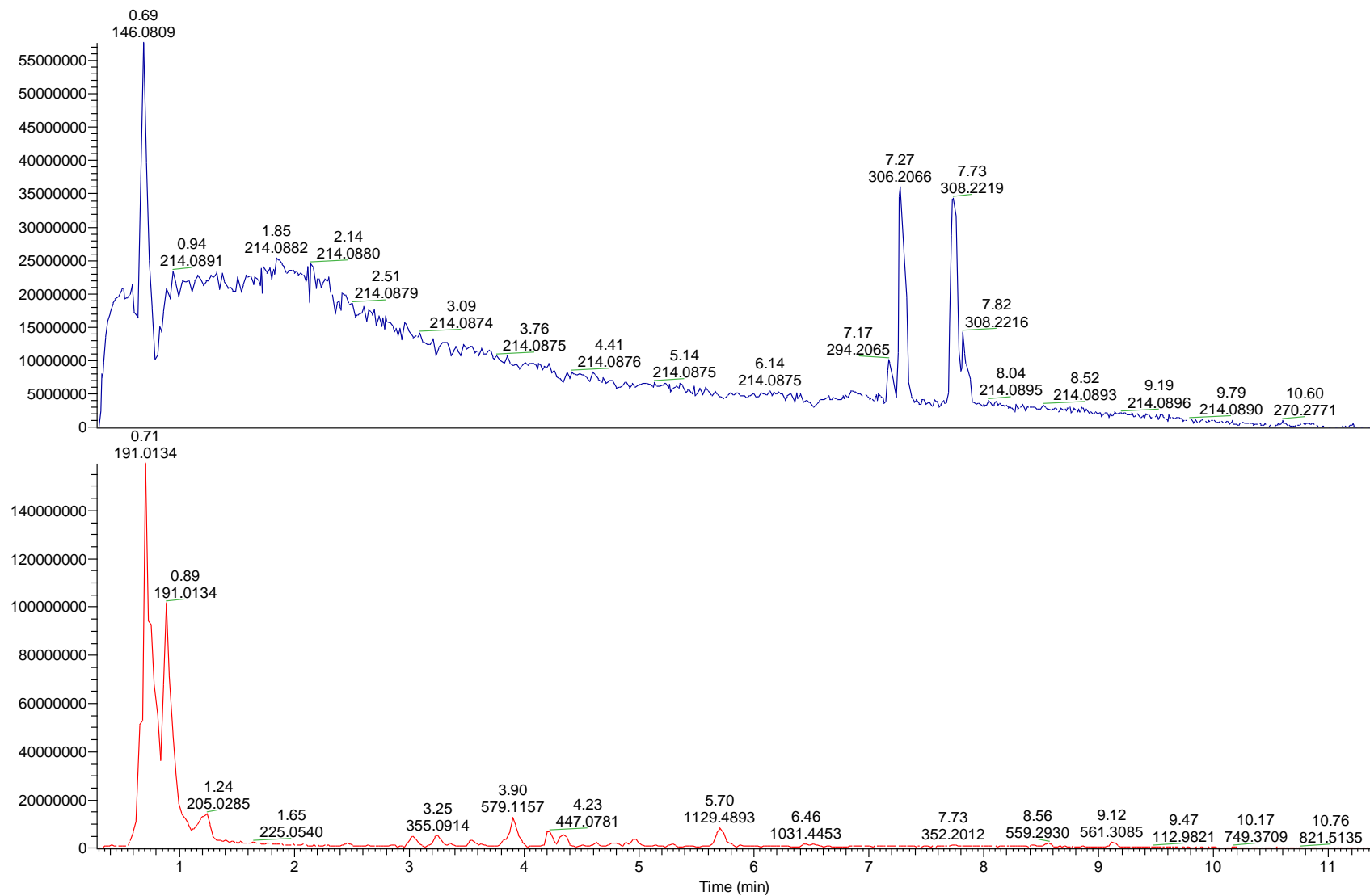


Figura 26. Perfil cromatográfico UHPLC-HRMS de los modos de ionización ESI (+, -) de la accesión Tomatito rojo (pico base)

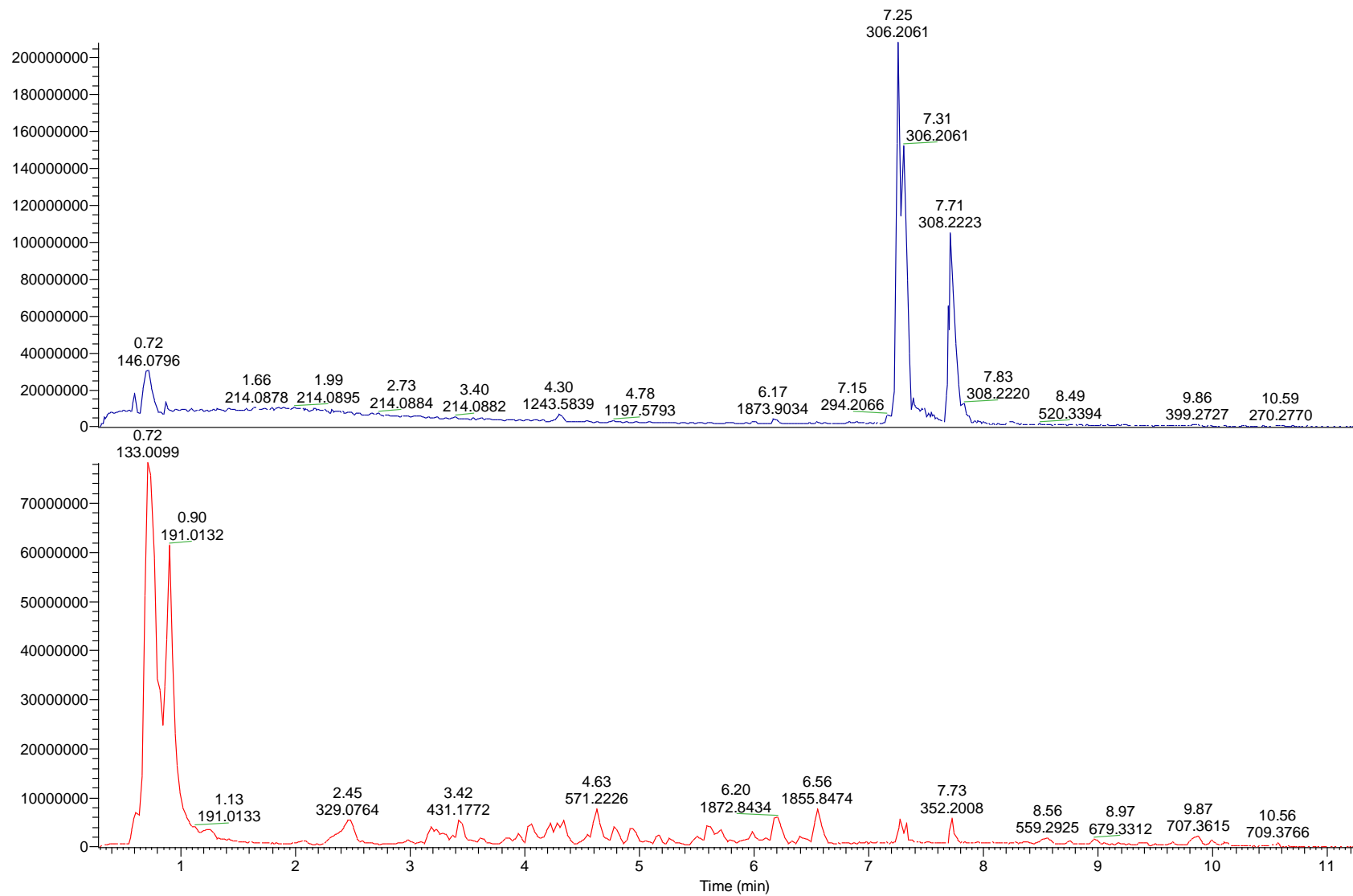


Figura 27. Perfil cromatográfico UHPLC-HRMS de los modos de ionización ESI (+, -) de la accesión Charapita (pico base)

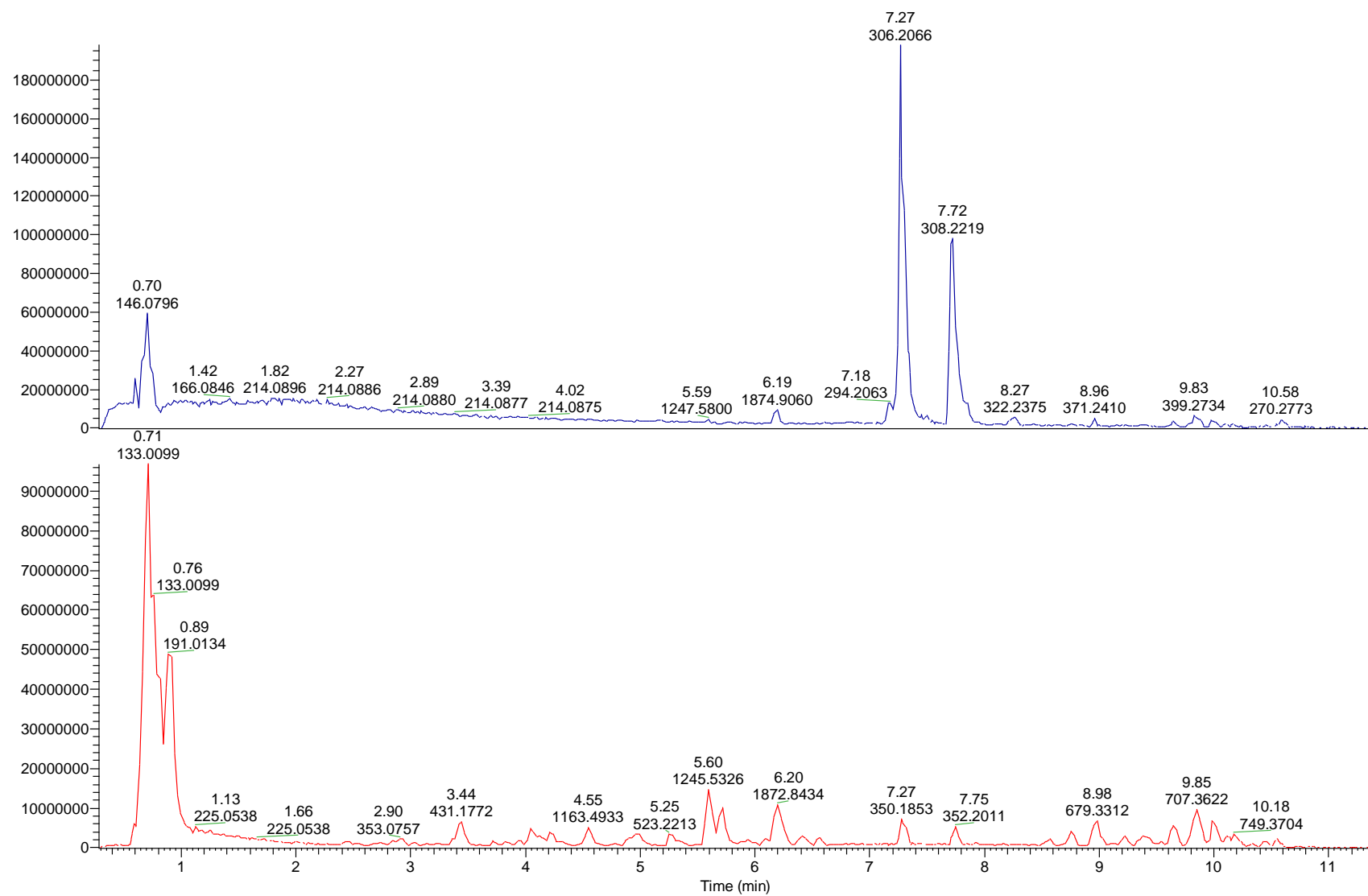


Figura 28. Perfil cromatográfico UHPLC-HRMS de los modos de ionización ESI (+, -) de la accesión Miscucho (pico base)

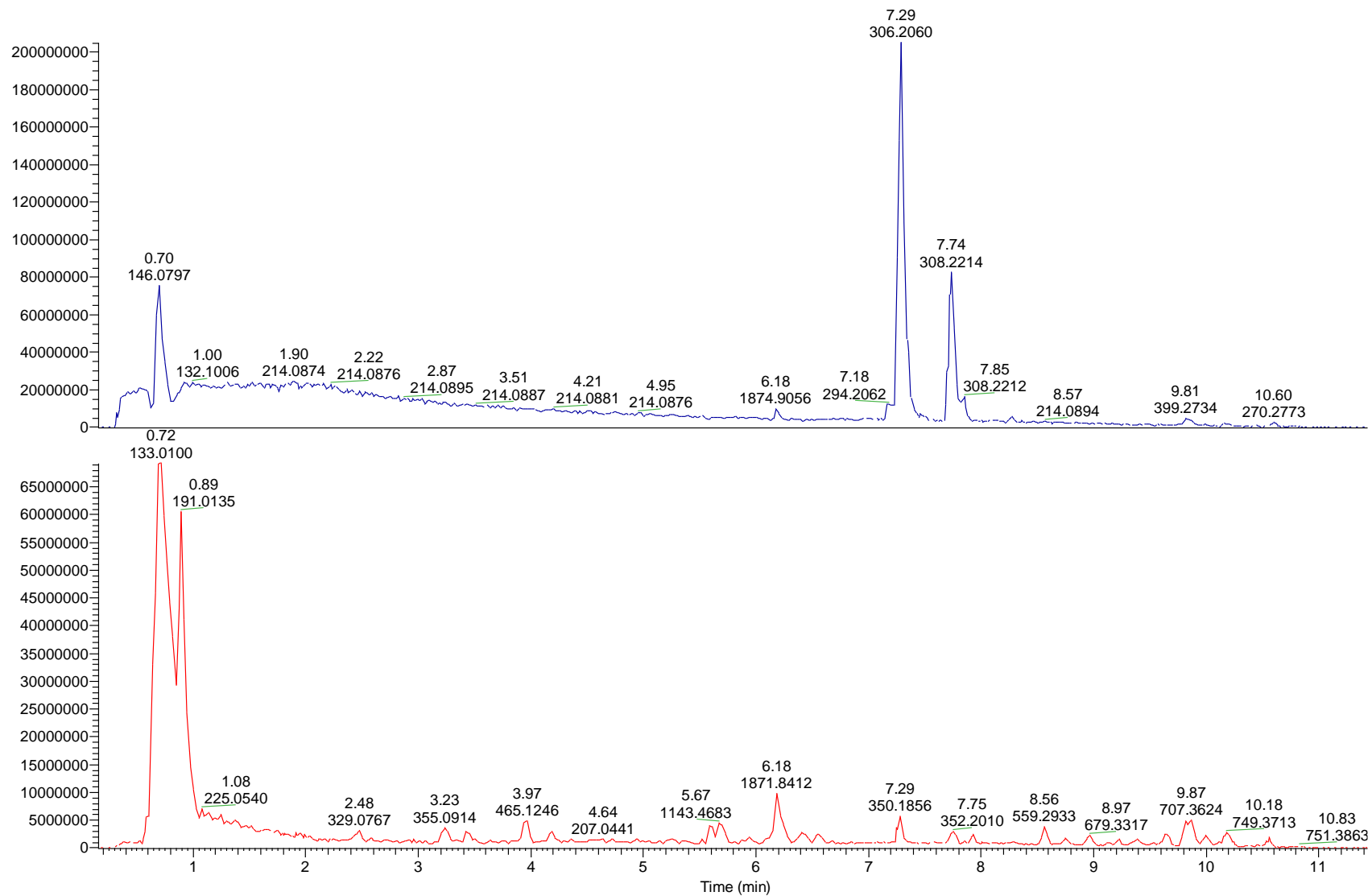


Figura 29. Perfil cromatográfico UHPLC-HRMS de los modos de ionización ESI (+, -) de la accesión Dulce rojo (pico base)

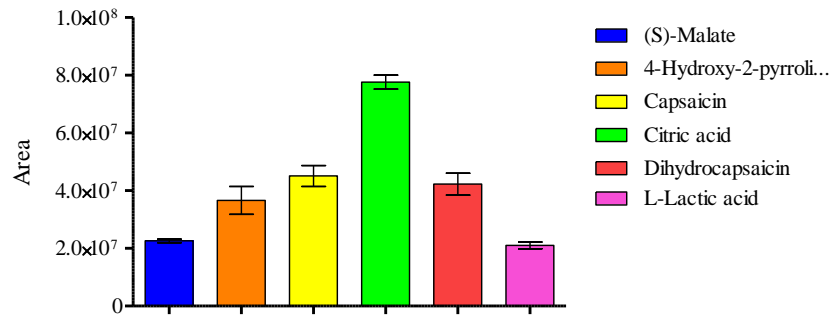


Figura 30. Diagrama de barras, promedio de áreas relativas (\pm SD) de los seis metabolitos mayoritarios del aji Ayuello.

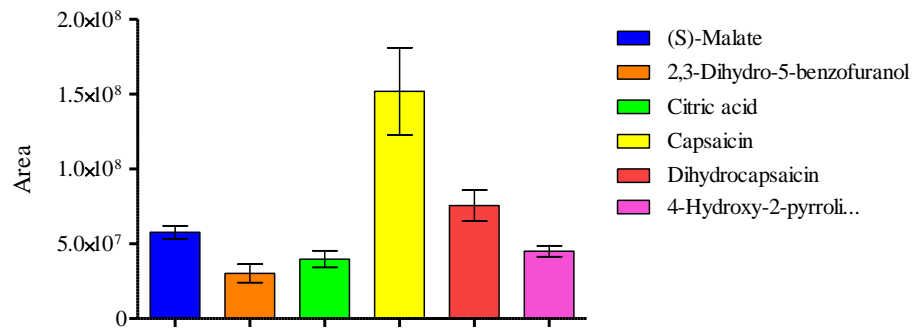


Figura 31. Diagrama de barras, promedio de áreas relativas (\pm SD) de los seis metabolitos mayoritarios del aji Charapita.

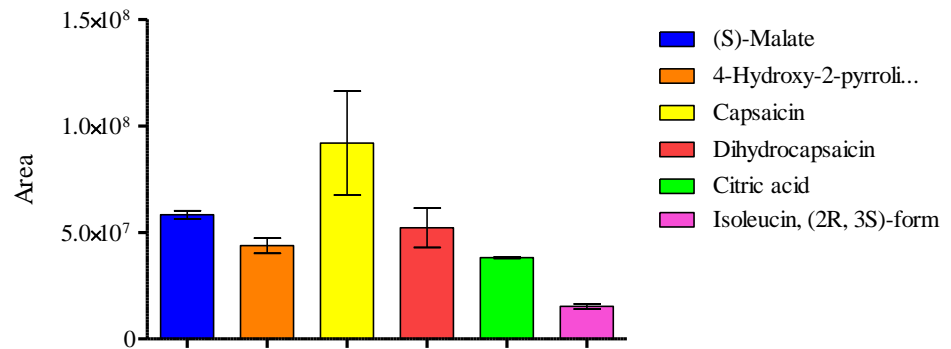


Figura 32. Diagrama de barras, promedio de áreas relativas (\pm SD) de los seis metabolitos mayoritarios del aji Dulce rojo.

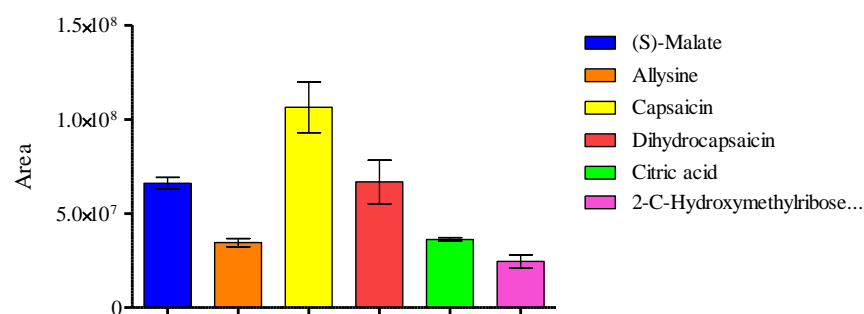


Figura 33. Diagrama de barras, promedio de áreas relativas (\pm SD) de los seis metabolitos mayoritarios del ají Miscucho.

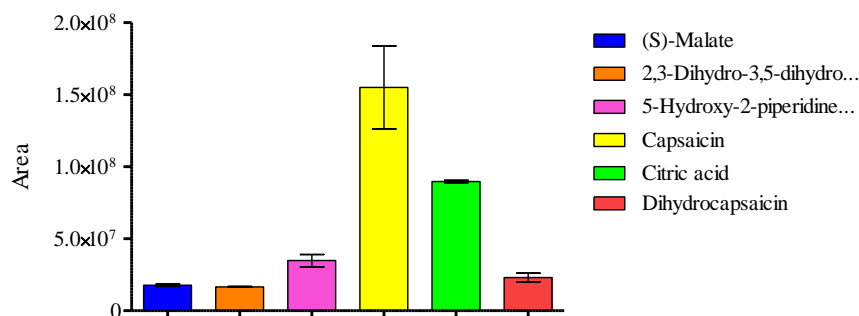


Figura 34. Diagrama de barras, promedio de áreas relativas (\pm SD) de los seis metabolitos mayoritarios del ají Tomatito rojo.

7.2. Perfil de metabolitos volátiles de los ajíes de Chinchá

Las figuras 35-39 muestran los resultados de los cromatogramas superpuestos utilizando 2 fibras SPME celeste y ploma en la extracción de metabolitos volátiles de los ajíes frescos y liofilizados, respectivamente. En ellos se observa que la fibra SPME celeste deja un tiempo de retención vacío (tiempo muerto) en el rango de 14-26 min que fue completado con los metabolitos extraídos de la fibra SPME ploma. Los dos cromatogramas fueron desarrollados utilizando una columna apolar DB5. Los metabolitos de la fibra SPME celeste son menos polares que los extraídos por la fibra ploma. La data cromatográfica obtenida por GC-MS de las muestras de ajíes

fresco y liofilizado fueron evaluadas con el software MS-DIAL para la alineación, deconvolución e identificación de los metabolitos volátiles. Las condiciones de trabajo por MS-DIAL versión 3.96 fueron las siguientes: 1) colección de data, masa inicial= 45 Da, masa final = 300 Da, tiempo de retención inicial = 8.09 min, tiempo de retención final = 68.33 min. 2) En la detección de características se utilizó una amplitud = 138802.6596 (fibra celeste) y amplitud = 31797.56614 (fibra ploma) que fueron calculados por experimentos en MS-DIAL para obtener aproximadamente 500 características. La identificación de los metabolitos volátiles se basó en el cálculo del índice de retención o Kovats. Para ello, se corrió una mezcla de alcanos (mezcla C8-C20) bajo las mismas condiciones cromatográficas que las muestras, con la finalidad de obtener sus respectivos tiempos de retención. La información de los tiempos de retención fue extraída utilizando el software AMDIS de NIST MS search 2.2. Con los datos obtenidos se elaboró una tabla de diccionario de alcanos. Esta tabla fue diseñada con los datos del tiempo de retención de los picos cromatográficos de los estándares de alcanos versus número de carbonos. La tabla fue guardada dentro de un archivo de texto (*.TXT), luego fue utilizada en el software MS-DIAL.

La alineación de los cromatogramas fue realizada en función del índice de retención (IR) con una tolerancia de 20 y tolerancia EI del 80%, con remoción de características basadas en la información del blanco.

Para la identificación de las características se utilizaron las mayores bases de datos de metabolitos volátiles: flavor2.0, mainlib y replib (NIST 14) y All public (MS-DIAL). En la identificación también se utilizaron 7 compuestos estándares de los que se identificaron los siguientes metabolitos: 3-carene, benzaldehyde, citral, citronellal, eucaliptol, menthone y R- (+)-limonene. De la deconvolución de la data cromatográfica de las 5 accesiones de ajíes, el número de metabolitos volátiles identificados en ají fresco fueron 237, mientras que para los ajíes liofilizados fueron 212 metabolitos.

En las tablas 67-86 se reportan las 40 principales características identificadas para las 5 accesiones de los ajíes frescos y liofilizados de Chincha luego de la alineación por MS-DIAL (ANEXO 2).

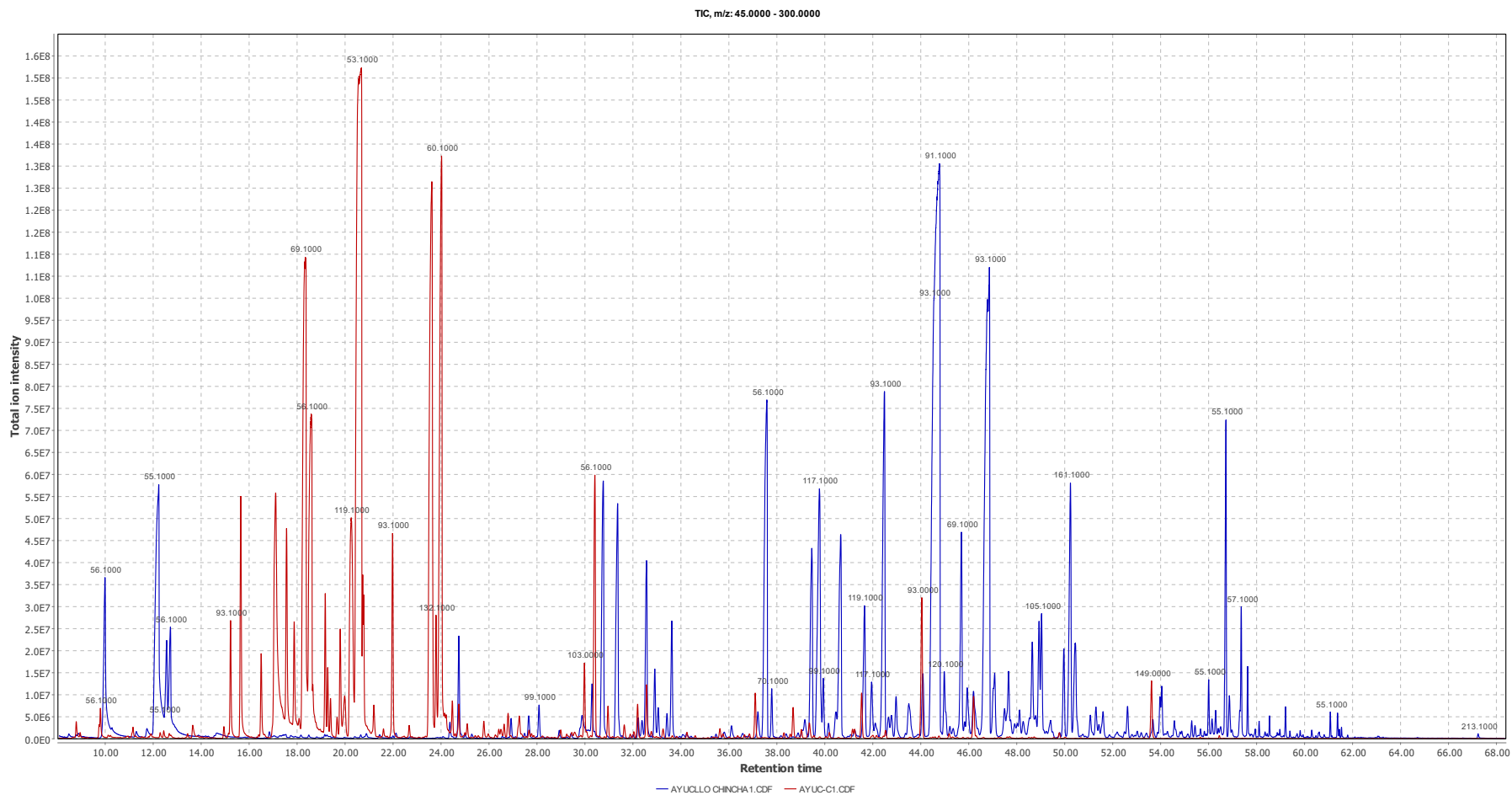


Figura 35. Superposición de cromatogramas de metabolitos volátiles obtenidos por SPME-GC-MS. En el cromatograma en rojo se observa los metabolitos del ají liofilizado (fibra ploma) y el cromatograma en azul se observa los metabolitos del ají fresco (fibra celeste) de la accesión Ayuclo.

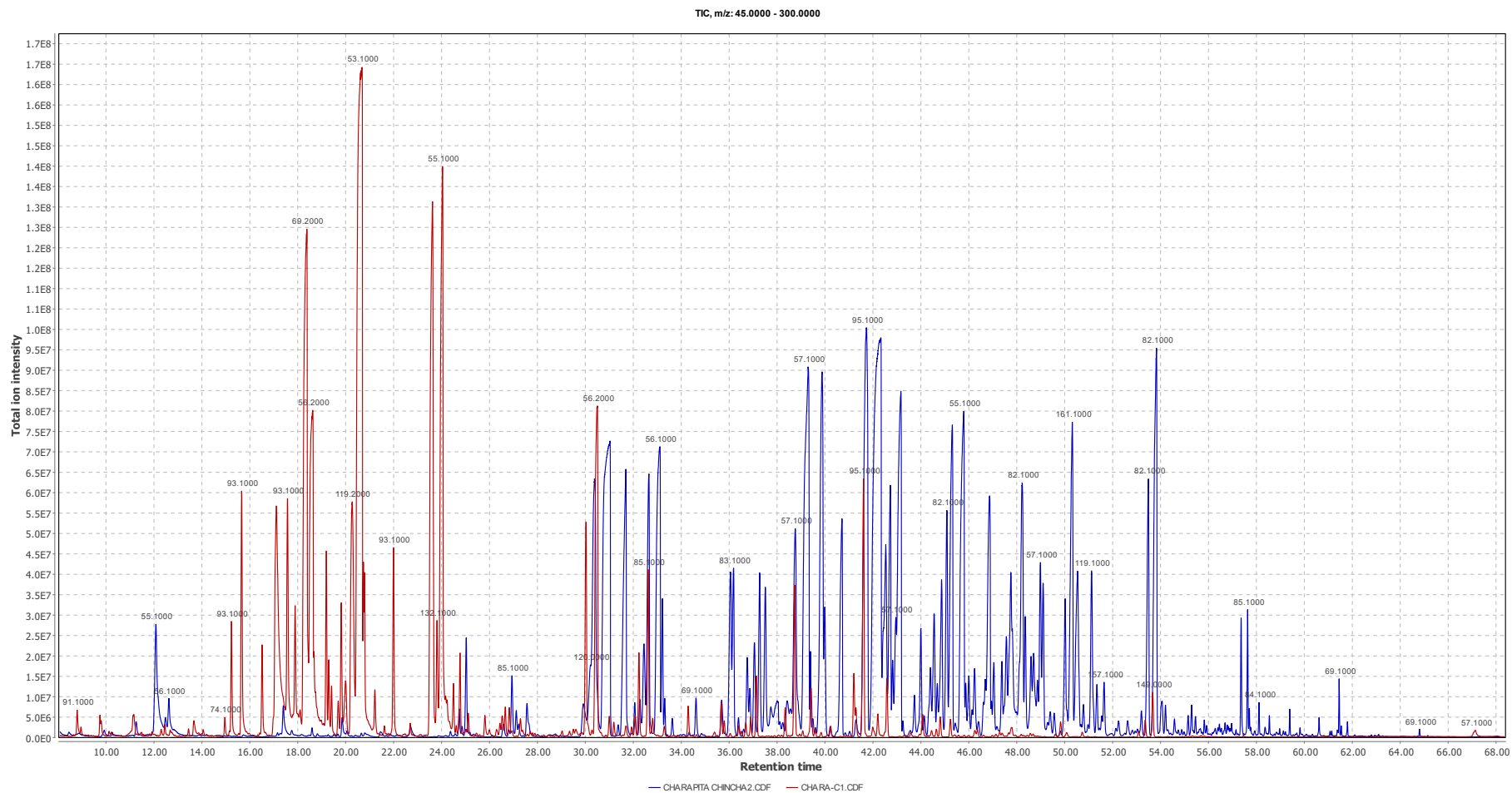


Figura 36. Superposición de cromatogramas de metabolitos volátiles obtenidos por SPME-GC-MS. En el cromatograma en rojo se observa los metabolitos del ají liofilizado (fibra ploma) y el cromatograma en azul se observa los metabolitos del ají fresco (fibra celeste) de la accesión Charapita.

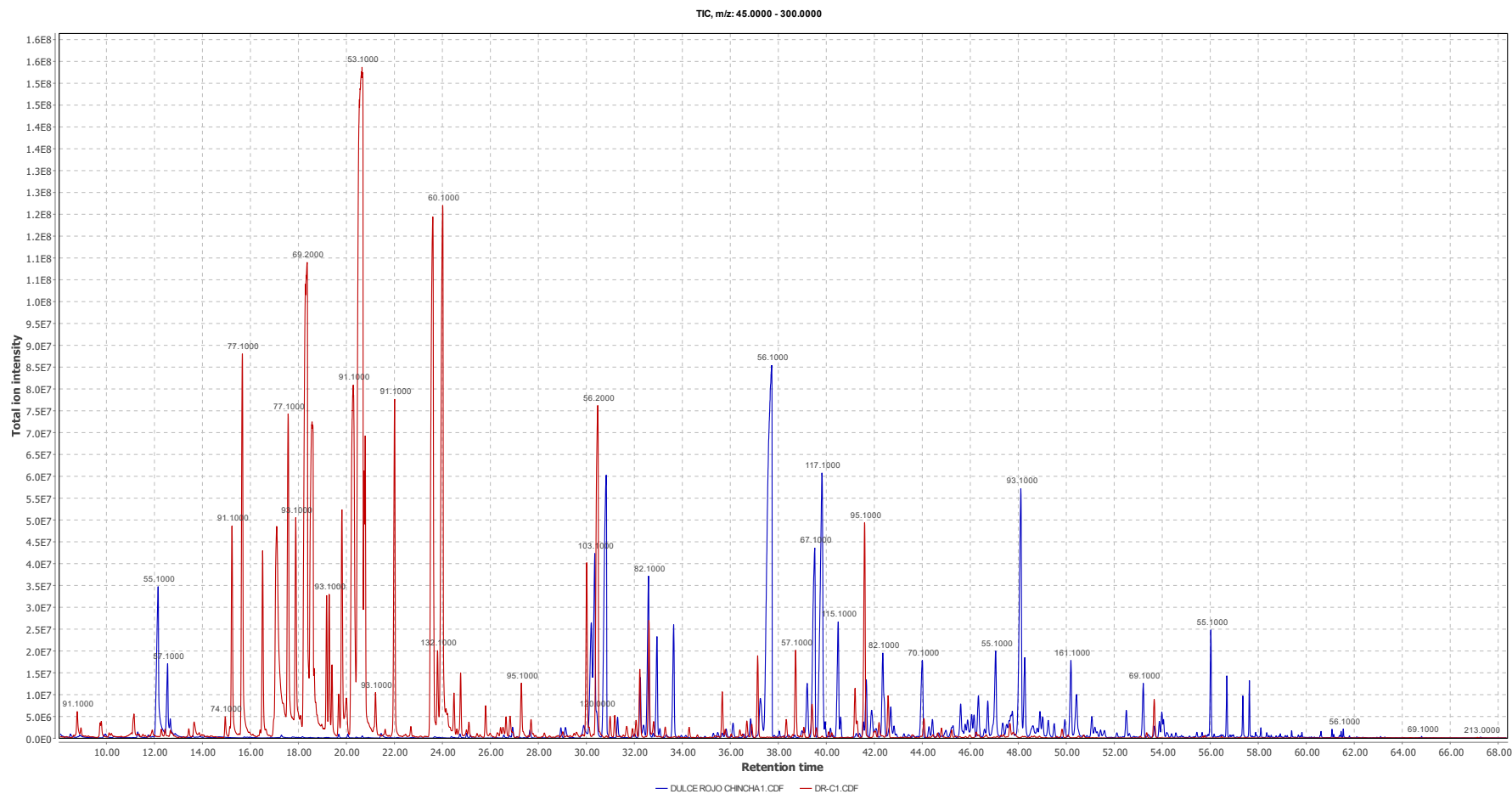


Figura 37. Superposición de cromatogramas de metabolitos volátiles obtenidos por SPME-GC-MS. En el cromatograma en rojo se observa los metabolitos del ají liofilizado (fibra ploma) y el cromatograma en azul se observa los metabolitos del ají fresco (fibra celeste) de la accesión Dulce rojo.

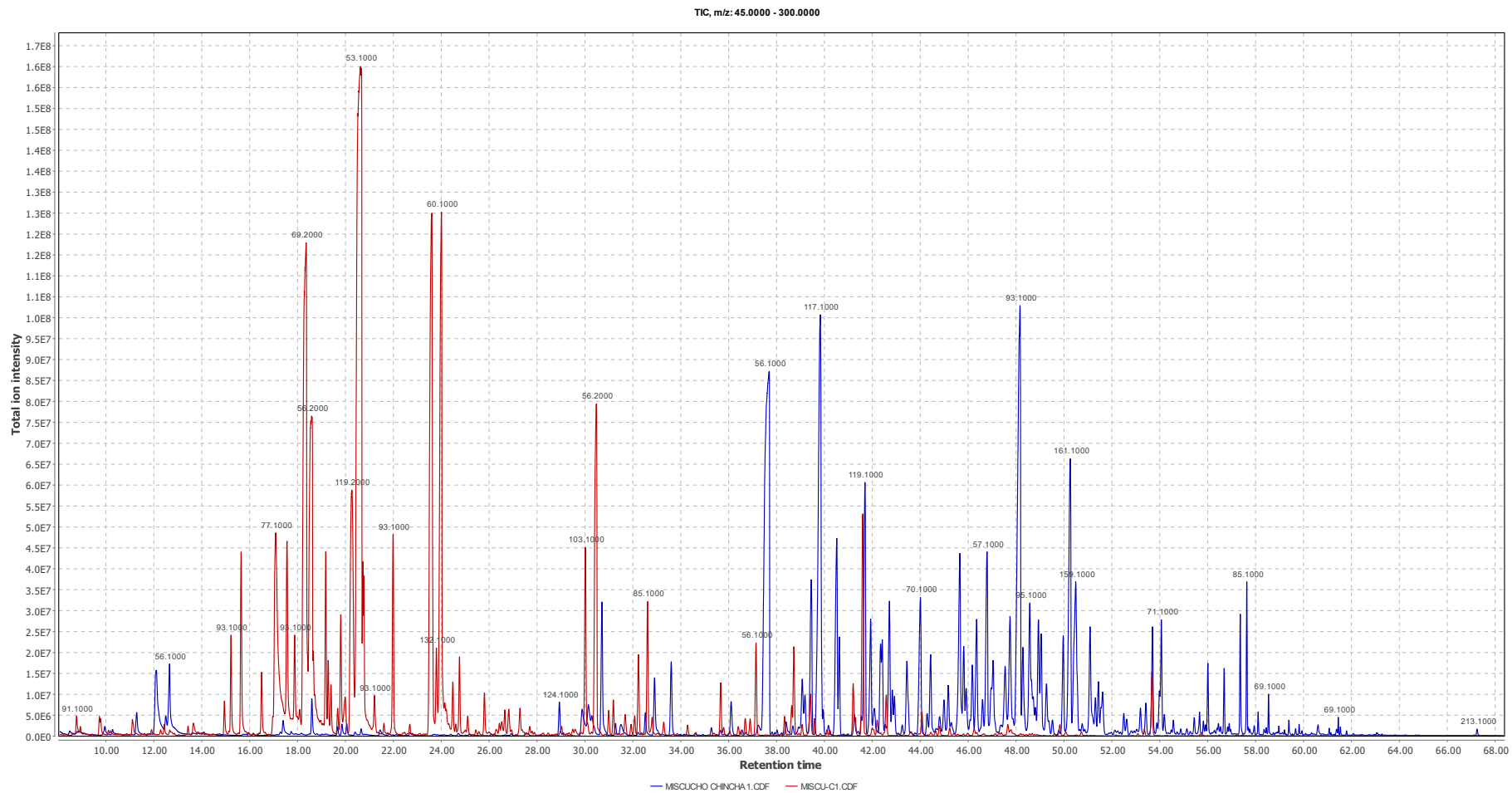


Figura 38. Superposición de cromatogramas de metabolitos volátiles obtenidos por SPME-GC-MS. En el cromatograma en rojo se observa los metabolitos del ají liofilizado (fibra ploma) y el cromatograma en azul se observa los metabolitos del ají fresco (fibra celeste) de la accesión Miscucho.

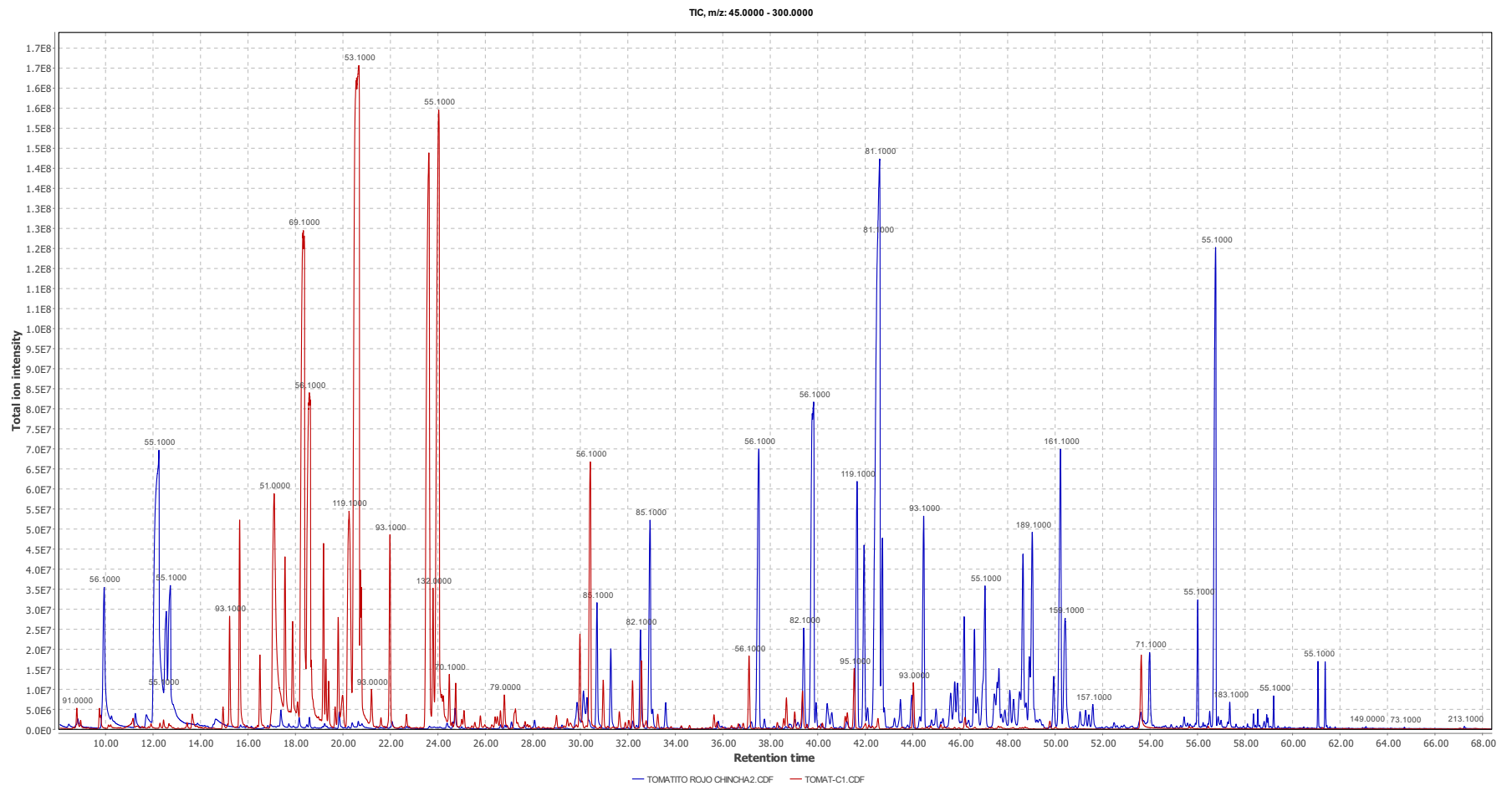


Figura 39. Superposición de cromatogramas de metabolitos volátiles obtenidos por SPME-GC-MS. En el cromatograma en rojo se observa los metabolitos del ají liofilizado (fibra ploma) y el cromatograma en azul se observa los metabolitos del ají fresco (fibra celeste) de la accesión Tomatito rojo.

7.3.Tratamiento de la data sensorial

7.3.1. Evaluación de la Normalidad de la data

La data sensorial fue sometida a los test de Shapiro Wilk, Anderson-Darling, lilliefors y Jarque-Bera (software XLSTAT de Excel) para rechazar o aceptar la hipótesis nula de que los datos (27 repeticiones por cada descriptor de una accesión dada) provienen de una población con distribución normal. Las pruebas de hipótesis para los test de normalidad son las siguientes: Si $p\text{-value} > 0.05$, se acepta H_0 = la variable donde proviene la data sigue una ley normal (hipótesis nula) y H_a = la variable donde proviene la data no sigue una ley normal (hipótesis alterna). Las tablas 5-9 muestran los resultados del test de normalidad para los descriptores de los ajíes de Chincha de data sensorial cruda. En rojo se indica las variables para los cuales se acepta la hipótesis alterna. Las tablas 10-14 muestran los resultados de test de normalidad de la data sensorial optimizada. La evaluación permitió juzgar sobre la representatividad de la data. Se observa que en general la data sensorial cruda proviene de una distribución normal.

Tabla 5. Test de Normalidad para la data sensorial del ají Ayuclo de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.025	0.047	0.271	0.311
Hierba	0.223	0.332	0.185	0.478
Manzana	0.407	0.621	0.767	0.589
Cítrico	0.367	0.273	0.060	0.501
Frutal	0.068	0.078	0.132	0.362
Maracuya	0.010	0.021	0.042	0.033
Acidez	0.160	0.195	0.189	0.461
Dulce	0.291	0.408	0.822	0.481
Pimiento	0.191	0.287	0.267	0.385
Tomate	0.325	0.392	0.439	0.593

Tabla 6. Test de Normalidad para la data sensorial del ají Charapita de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.010	0.012	0.023	0.380
Hierba	0.093	0.139	0.333	0.503
Manzana	0.061	0.064	0.005	0.452
Cítrico	0.342	0.183	0.144	0.523
Frutal	0.013	0.017	0.055	0.273
Maracuya	0.016	0.019	0.049	0.287
Acidez	0.401	0.477	0.462	0.653
Dulce	0.000	< 0.0001	0.005	0.055
Pimiento	0.014	0.023	0.072	0.269
Tomate	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Tabla 7. Test de Normalidad para la data sensorial del ají Dulce Rojo de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.025	0.058	0.196	0.210
Hierba	0.046	0.054	0.254	0.385
Manzana	0.083	0.124	0.267	0.379
Cítrico	0.385	0.535	0.582	0.604
Frutal	0.390	0.530	0.716	0.549
Maracuya	0.005	0.034	0.100	0.005
Acidez	0.010	0.028	0.104	0.185
Dulce	0.020	0.020	0.129	0.381
Pimiento	0.430	0.569	0.522	0.586
Tomate	0.086	0.101	0.070	0.420

Tabla 8. Test de Normalidad para la data sensorial del ají Miscucho de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.003	0.006	0.091	0.102
Hierba	0.097	0.070	0.037	0.449
Manzana	0.060	0.069	0.024	0.486
Cítrico	0.093	0.075	0.063	0.366
Frutal	0.055	0.058	0.131	0.358
Maracuya	0.010	0.018	0.052	0.142
Acidez	0.026	0.048	0.322	0.300
Dulce	0.136	0.230	0.398	0.479
Pimiento	0.417	0.693	0.855	0.694
Tomate	0.012	0.014	0.055	0.120

Tabla 9. Test de Normalidad para la data sensorial del ají Tomatito rojo de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.000	< 0.0001	0.002	0.004
Hierba	0.003	0.001	0.025	0.162
Manzana	0.010	0.013	0.073	0.250
Cítrico	0.007	0.005	0.017	0.219
Frutal	0.068	0.135	0.504	0.377
Maracuya	< 0.0001	< 0.0001	0.002	0.001
Acidez	0.001	0.003	0.011	0.001
Dulce	0.014	0.010	0.023	0.348
Pimiento	0.195	0.383	0.585	0.399
Tomate	0.044	0.075	0.228	0.274

Tabla 10. Test de Normalidad para la data sensorial optimizada del ají Ayuello de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.684	0.692	0.885	0.935
Hierba	0.507	0.507	0.872	0.624
Manzana	0.209	0.282	0.056	0.129
Cítrico	0.234	0.478	0.696	0.332
Frutal	0.389	0.622	0.685	0.133
Maracuya	0.983	0.967	0.954	0.884
Acidez	0.409	0.312	0.621	0.383
Dulce	0.680	0.592	0.610	0.657
Pimiento	0.122	0.290	0.724	0.229
Tomate	0.743	0.523	0.720	0.897

Tabla 11. Test de Normalidad para la data sensorial optimizada del ají Charapita de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.438	0.700	0.401	0.365
Hierba	0.797	0.865	0.974	0.821
Manzana	0.490	0.508	0.745	0.293
Cítrico	0.196	0.289	0.383	0.068
Frutal	0.968	0.841	0.548	0.911
Maracuya	0.738	0.786	0.904	0.754
Acidez	0.888	0.810	0.831	0.746
Dulce	0.498	0.218	0.224	0.828
Pimiento	0.597	0.502	0.449	0.880
Tomate	0.904	0.959	0.987	0.524

Tabla 12. Test de Normalidad para la data sensorial optimizada del ají Dulce rojo de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.605	0.606	0.802	0.188
Hierba	0.802	0.524	0.538	0.785
Manzana	0.228	0.323	0.469	0.166
Cítrico	0.737	0.732	0.797	0.789
Frutal	0.286	0.519	0.576	0.361
Maracuya	0.305	0.322	0.373	0.398
Acidez	0.817	0.518	0.585	0.722
Dulce	0.580	0.799	0.780	0.722
Pimiento	0.730	0.335	0.312	0.752
Tomate	0.888	0.651	0.310	0.869

Tabla 13. Test de Normalidad para la data sensorial optimizada del ají Miscucho de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.316	0.288	0.261	0.435
Hierba	0.930	0.902	0.882	0.861
Manzana	0.688	0.540	0.525	0.460
Cítrico	0.170	0.273	0.165	0.219
Frutal	0.927	0.948	0.855	0.625
Maracuya	0.152	0.300	0.381	0.656
Acidez	0.709	0.857	0.733	0.847
Dulce	0.836	0.901	0.899	0.828
Pimiento	0.250	0.055	0.064	0.356
Tomate	0.719	0.561	0.494	0.733

Tabla 14. Test de Normalidad para la data sensorial optimizada del ají Tomatito rojo de Chincha

Variable\Test	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
Orégano	0.603	0.505	0.601	0.786
Hierba	0.885	0.965	0.924	0.820
Manzana	0.767	0.718	0.640	0.504
Cítrico	0.717	0.454	0.314	0.319
Frutal	0.093	0.082	0.231	0.158
Maracuya	0.880	0.674	0.258	0.685
Acidez	0.495	0.682	0.815	0.886
Dulce	0.547	0.526	0.849	0.831
Pimiento	0.972	0.903	0.601	0.994
Tomate	0.490	0.276	0.346	0.870

7.3.2. Análisis del panel de degustación

El análisis de Panel de la data sensorial cruda permitió evaluar la capacidad del panel para diferenciar de manera reproducible las notas del sabor y aroma, su confiabilidad, su repetibilidad e identificar grupos homogéneos entre los catadores. En la primera etapa, se lleva a cabo una serie de ANOVA, donde el objetivo es verificar para cada descriptor si se pone en evidencia un efecto sobre el producto. La tabla 15 muestra el resumen del análisis de serie de ANOVA en la cual se compararon los *p*-value asociados al factor producido por los diferentes descriptores.

Tabla 15. Identificación de descriptores por el cual hay un efecto del producto (*p* Value <0.05). Los valores en negrita corresponden a los descriptores eliminados

Descripteurs	F	Pr > F
Orégano	1.806	0.036
Hierba	1.385	0.157
Manzana	1.744	0.045
Cítrico	5.912	< 0.0001
Frutal	6.445	< 0.0001
Maracuya	3.258	< 0.0001
Acidez	2.945	0.000
Dulce	11.549	< 0.0001
Pimiento	14.967	< 0.0001
Tomate	12.146	< 0.0001

En la figura 40 se observa el poder de discriminación de cada catador por el cual ha podido diferenciar los productos en unidades de porcentaje.

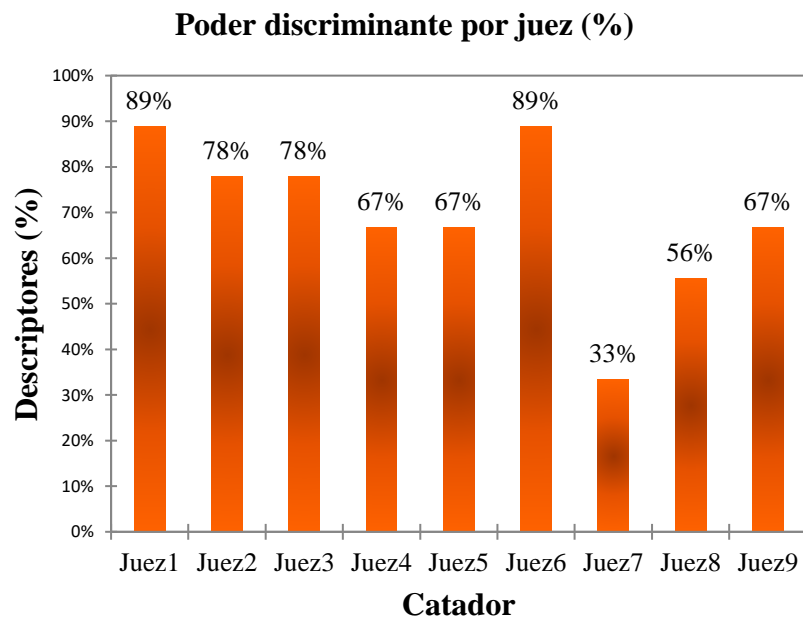


Figura 40. Capacidad de discriminación de los Panelistas (%)

7.3.3. Análisis proximal e instrumental

En la tabla 16 se muestran los resultados del análisis proximal e instrumental, con tres réplicas en cada determinación y una cuarta réplica fue tomada como el promedio de las tres. Los resultados fueron utilizados para realizar la correlación con las variables de la data sensorial optimizada (ver sección 15.3.8)

Tabla 16. Resultados de las determinaciones de análisis proximal e instrumental

Accesión	Especie	% humedad	% grasa	Proteínas	pH	% Brix
Ayucllo	Capbac	82.46	12.86	7.95	5.01	1.6
Ayucllo	Capbac	82.27	12.90	7.44	5.02	1.6
Ayucllo	Capbac	84.18	12.80	7.35	5.04	1.6
Ayucllo	Capbac	82.97	12.85	7.58	5.02	1.6
Charapita	Capfrut	80.32	7.84	11.20	5.19	1.2
Charapita	Capfrut	79.72	7.88	10.74	5.24	1.2
Charapita	Capfrut	80.45	7.75	10.88	5.2	1.2
Charapita	Capfrut	80.16	7.83	10.94	5.21	1.2
Dulce R	Capchin	89.08	2.40	9.66	5.14	2.1
Dulce R	Capchin	89.05	2.40	9.47	5.17	2.1
Dulce R	Capchin	89.37	2.41	9.43	5.12	2.1
Dulce R	Capchin	89.17	2.41	9.52	5.14	2.1
Miscucho	Capchin	88.94	4.18	8.88	5.3	1.9
Miscucho	Capchin	88.48	4.19	8.53	5.27	1.9
Miscucho	Capchin	88.40	4.19	8.70	5.28	1.9
Miscucho	Capchin	88.61	4.19	8.70	5.28	1.9
Tomatito R	Capbac	81.74	3.58	3.98	5.03	1.7
Tomatito R	Capbac	80.49	3.64	4.37	4.9	1.7
Tomatito R	Capbac	82.11	3.58	4.12	4.89	1.7
Tomatito R	Capbac	78.35	3.60	4.15	4.94	1.7

7.3.4. Optimización del modelo del perfil sensorial

La optimización del modelo de perfil sensorial multivariado fue llevada a cabo utilizando los parámetros de la media y la desviación estándar de la data sensorial original. Esta información fue introducida en el software (Autograph v.4.0.12, 30-day trial) para generar una población de $N = 300$ datos con distribución normal de cada descriptor sensorial. En la figura 42B se observa el modelo OPLS-DA resultante.

7.3.5. Evaluación de la calidad del modelo sensorial optimizado

El modelo sensorial fue evaluado mediante análisis multivariado (PCA y OPLS-DA). La calidad del modelo sensorial OPLS-DA fue juzgada con los parámetros de calidad Q^2 y R^2 . El valor de R^2 indica qué tan bien el modelo explica el conjunto de datos, y la correlación con validación cruzada Q^2 describe la bondad de la predicción, (Q^2 es la fracción de la variación total de las variables que pueden predecir un componente principal utilizando el método CV (validación cruzada)). La validación cruzada es un método que implica dividir a los sujetos o individuos en subconjuntos y ajustar el modelo después de excluir aleatoriamente un subconjunto a la vez en el análisis. R^2 no debe ser considerablemente más alto que Q^2 . La diferencia entre los valores R^2 y Q^2 proporciona información confiable sobre la bondad de ajuste del modelo. Si la diferencia excede de 0.2 a 0.3, implica un sobreajuste, lo que significa que la robustez del modelo es pobre (111). Antes de

ajustar los modelos supervisados, se realizó un análisis preliminar de los componentes principales para una visión general de los datos, detectando valores atípicos y grupos entre las observaciones.

Los valores atípicos fuertes fueron identificados a partir de los scores y los gráficos en el rango T2 de Hotelling. Esta última es una generalización multivariada de la prueba t de Student, que proporciona un control de observación que se adhiere a la normalidad multivariada. Cuando se usa junto con una gráfica de scores, la T2 de Hotelling define el área de normalidad correspondiente al 95% de nivel de confianza.

La prueba de permutación de respuesta (con $N = 100$) se utilizó para validar la capacidad predictiva de los modelos OPLS-DA. En esta prueba estadística, los valores R^2Y y Q^2Y se calcularon para los modelos permutados y se compararon con los valores de los modelos verdaderos. La prueba se lleva a cabo asignando aleatoriamente a dos grupos diferentes, después de lo cual los modelos OPLS-DA se ajustan a cada variable de clase permutada. Los resultados permitieron juzgar sobre la calidad del modelo. Si los valores de R^2Y y Q^2Y tienen valores menores que los modelos calculados reales, se dice que el modelo verdadero OPLS-DA calculado es estadísticamente mucho mejor que los 100 modelos permutados para cada conjunto de datos.

Para validar aún más el modelo OPLS-DA se utilizó el análisis de varianza de validación cruzada (CV-ANOVA). Para los métodos supervisados como OPLS-DA, el valor p ANOVA (CV-ANOVA) con validación cruzada puede ser calculado como una medida de significancia para la separación observada del grupo. El valor del CV-ANOVA se basa en su capacidad de proporcionar una prueba de significancia formal de la misma forma que el ANOVA ordinario. Se consideró significativo para un $p < 0.05$.

Adicionalmente, también se realizó los análisis de las curvas de operador receptor (ROC) para validar la robustez y bondad del modelo OPLS-DA. La curva ROC fue generada a partir de la validación cruzada siete veces obtenida por SIMCA-P 15. Para cada modelo OPLS-DA se calculó el área bajo una curva (ROC). El área bajo la curva ROC (AUC) es una medida del rendimiento del modelo de clasificación. El valor de esta área puede interpretarse como la probabilidad de que, dado un par de clases, una que pertenezca a una clase y la otra no, luego el modelo clasifica correctamente ambas clases.

7.3.6. Estimación de la confiabilidad del modelo sensorial

La figura 41 muestra el AMV de la data sensorial cruda (PCA y OPLS-DA). Se construyó el modelo de PCA para tres variables categóricas, las tres especies de ajíes C-bac= *C. baccatum*, C-chin= *C. chinense* y C-frut = *C. frutescens*. El modelo dio un valor de R^2 igual a 0.715, sin embargo, gráficamente no se evidencia una

fuerte separación de las clases. Después del PCA, los conjuntos de datos fueron procesados utilizando un método estadístico supervisado más potente, OPLS-DA. Para validar la significancia estadística de cada modelo OPLS-DA, los valores de R^2Y y Q^2 fueron calculados en base a una validación cruzada (CV) de siete veces. La construcción del modelo OPLS-DA para la discriminación de las tres clases fue llevada a cabo por proyección de los individuos o muestras sobre el eje Y, este también es un modelo lineal. En el modelo OPLS-DA se observa un mejor agrupamiento categórico por especies, sin embargo, los valores de R^2Y y Q^2Y son menores a 0.5, lo que evidencia una pobre capacidad de ajuste, de discriminación y predicción. El PCA es un modelo lineal, y como tal, el valor de R^2 es una medida de la bondad o precisión de ajuste del modelo (para un modelo con tres variables categóricas). Sin embargo, el valor de R^2 no es una medida definitiva para juzgar la calidad del modelo. El parámetro R^2 describe la fracción de los datos originales explicados por el modelo (si $R^2 = 1.0$ significa que el modelo explica el 100% de los datos). Mientras que Q^2 describe la capacidad predictiva del modelo y es la fracción de los datos originales explicada por el modelo de validación cruzada que mide la capacidad del modelo para predecir un nuevo conjunto de datos (112). Se considera valores aceptables de R^2 y Q^2 mayores a 0.7 (SIMCA P considera un valor de $Q^2 > 0.5$ como un buen modelo predictivo) (113). La figura 42, muestra el AMV de la data sensorial optimizada. En el PCA se observa una clara tendencia de separación de clases. El valor de R^2 es igual a 0.72. El modelo OPLS-DA muestra valores de R^2Y y $Q^2Y > 0.7$, que son valores aceptables para la data sensorial optimizada revelando una alta capacidad de discriminación y predicción. La figura

43 muestra el análisis de permutaciones con 100 iteraciones para la validación de los modelos OPLS-DA de la data sensorial cruda y optimizada. Los valores de R^2 y Q^2 estimados de las permutaciones son menores que los valores reales. Los resultados indican que los modelos calculados tienen valores de R^2 y Q^2 mucho más altos y, por lo tanto, los verdaderos modelos OPLS-DA calculados son estadísticamente mucho mejores que los 100 modelos permutados para cada conjunto de datos (113,114).

La robustez del modelo OPLS-DA de la data sensorial optimizada fue validada aún más utilizando el área bajo la curva ROC (AUC). La figura 44 muestra los gráficos ROC de la data sensorial cruda y optimizada. El área bajo la curva ROC (AUC) es una medida del rendimiento del modelo de clasificación. Una AUC cercana a 1 indica un modelo de clasificación exitoso de ideal separabilidad. Mientras que un $AUC = 0.5$ significa que el modelo no tiene capacidad discriminante. En un típico gráfico ROC la sensibilidad o Tasa positiva verdadera (TPR) se grafica en el eje Y, y en el eje X se encuentra la tasa de falsos positivos ($FPR = 1 - \text{especificidad}$). La sensibilidad, especificidad y precisión fueron determinados sobre la base de la predicción de la clase de muestra durante la validación cruzada siete veces (Y-predcv) en el software SIMCA-P. En el caso de OPLS-DA de las tres clases de la data sensorial optimizada, las curvas (ROC) del modelo arrojaron una separación entre las clases individuales uno a uno con un área bajo la curva de $AUC(C-bac) = 1$, $AUC(C-frut) = 1$, $AUC(C-chin) = 0.989583$ (Figura 44B).

La figura 45 muestra los resultados de la validación externa del modelo OPLS-DA de la data sensorial optimizada. La figura 45A muestra el gráfico de score predichos a partir del modelo OPLS-DA de la data sensorial optimizada utilizando la data sensorial original. Se observa gráficamente una excelente separación de clases predichas. La figura 45B muestra el gráfico de barras del promedio de las probabilidades de las clases predichas. Se observa que el modelo OPLS-DA (data optimizada) predice nuevos scores cuyos valores predichos por el modelo tienen valores de probabilidad mayores a 0.5, lo que confirma que el modelo tiene una alta capacidad de predicción.

La tabla 17 muestra los resultados de Análisis de varianza con validación cruzada (CV-ANOVA) utilizado para evaluar la fiabilidad de los modelos obtenidos. El análisis proporciona el p -value que indica el nivel de significancia de la separación de las clases en el análisis discriminante del modelo OPLS-DA. Los modelos de regresión OPLS-DA de la data sensorial cruda y optimizada fueron validados por (CV-ANOVA), con un p -value < 0.05.

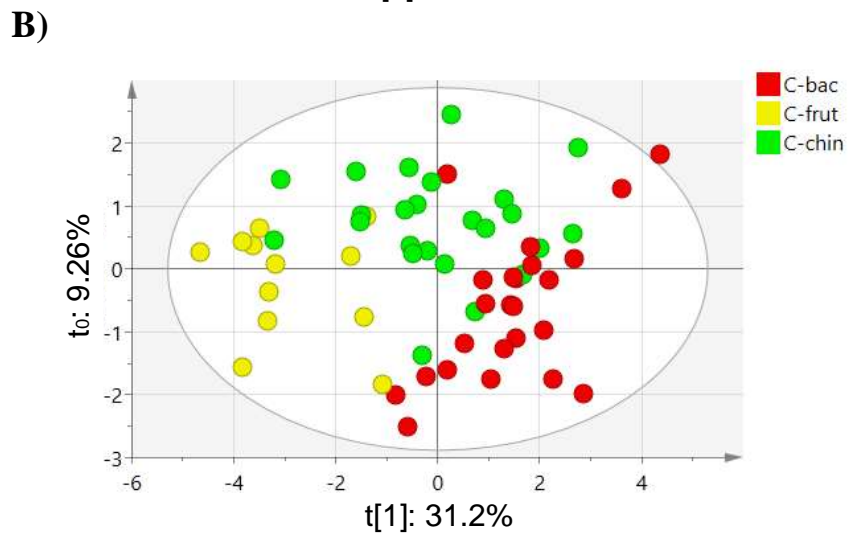
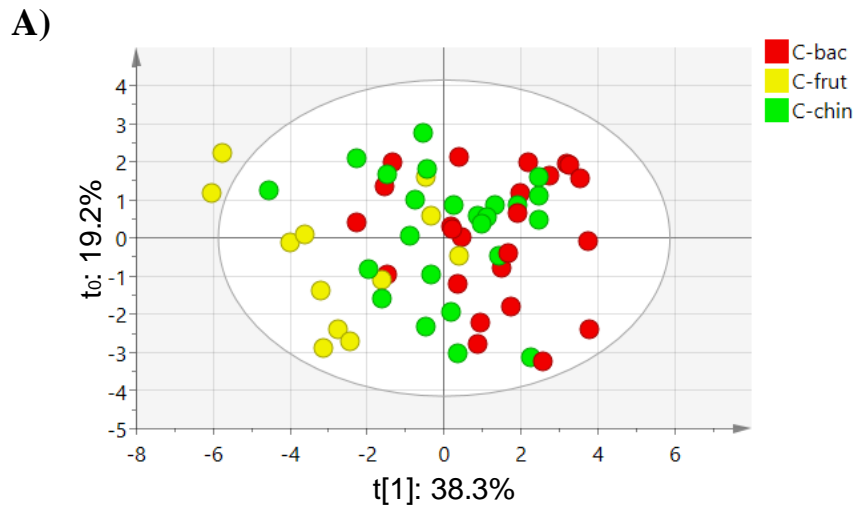
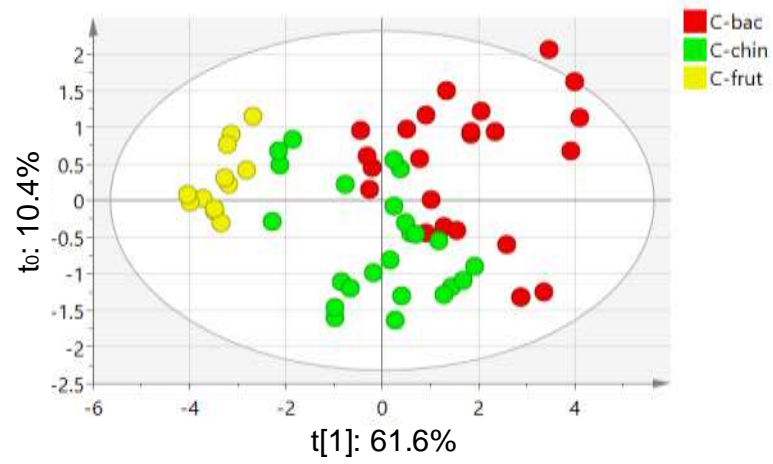


Figura 41. Análisis multivariado de la data sensorial original. En A se observa el modelo de PCA ($R^2X = 0.715$), en B se observa el modelo OPLS-DA ($R^2Y = 0.435$, $Q^2 = 0.31$). La varianza explicada es de 57.5 y 40.46%, respectivamente.

A)



B)

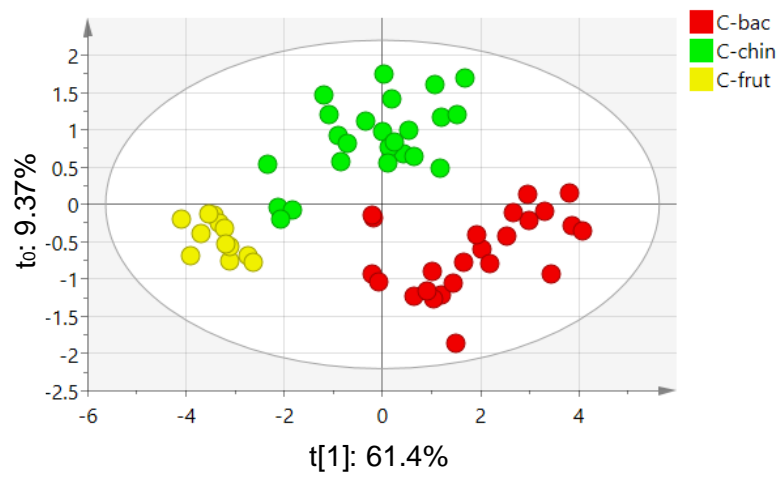
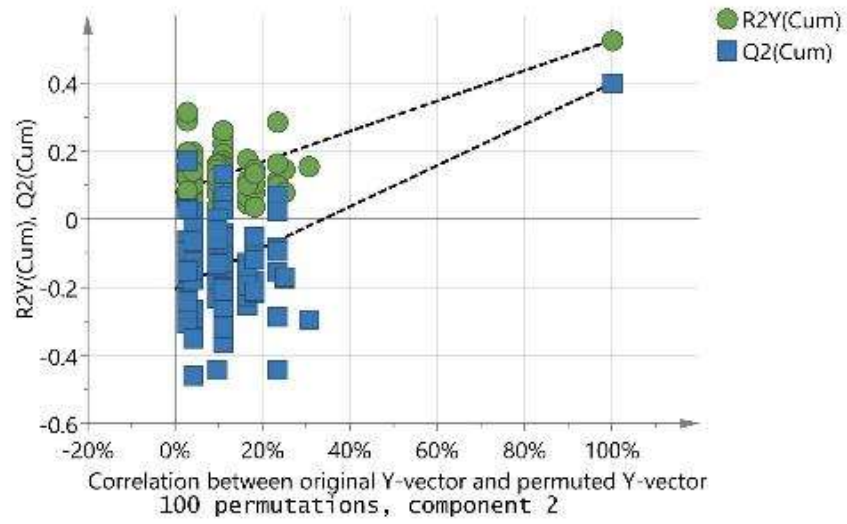


Figura 42. Análisis multivariado de la data sensorial optimizada. En A se observa el modelo de PCA ($R^2 = 0.72$), en B se observa el modelo OPLS-DA ($R^2Y = 0.73$, $Q^2 = 0.705$). La varianza explicada es de 72 y 70.77%, respectivamente.

A)



B)

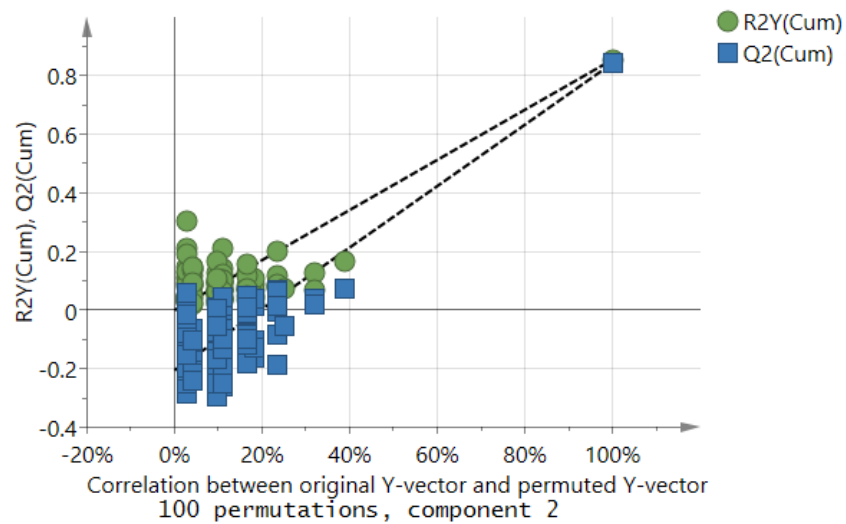
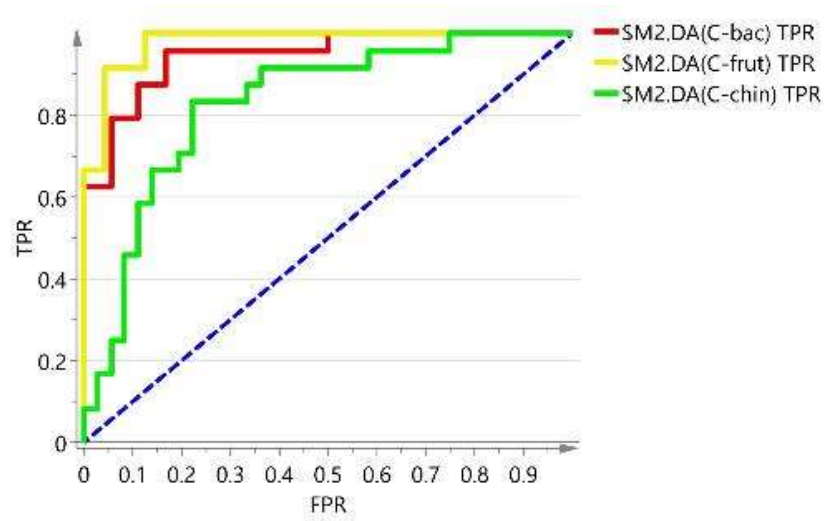


Figura 43. Prueba de permutación para la validación del modelo OPLS-DA la prueba de permutación de respuesta fue realizada con 100 repeticiones. En A se observa la prueba de permutación para el modelo OPLS-DA de la data sensorial real, interceptos: $R^2 = (0.0, 0.0798)$, $Q^2 = (0.0, -0.204)$. En B se observa la prueba de permutación de la data sensorial optimizada, interceptos: $R^2 = (0.0, -0.000444)$, $Q^2 = (0.0, -0.205)$.

A)



B)

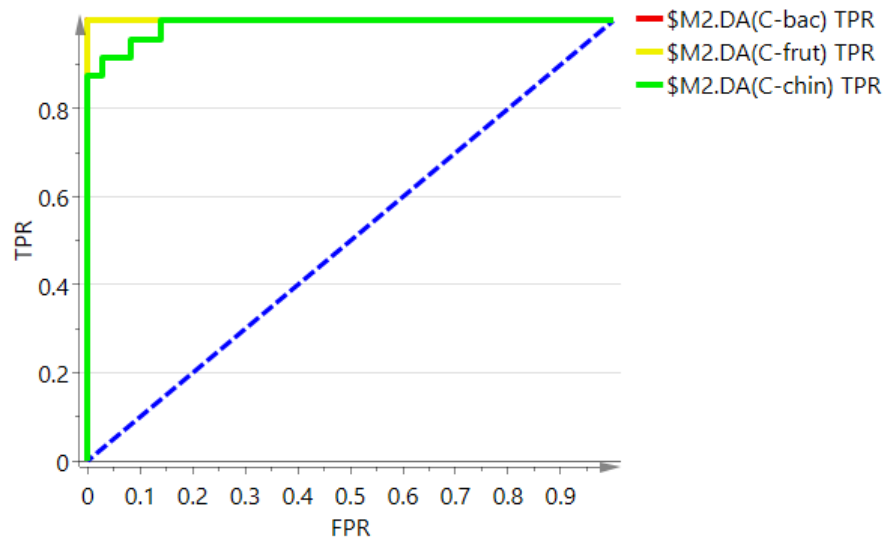


Figura 44. Gráficos ROC para la evaluación del rendimiento del clasificador OPLS-DA por especies de ajíes en función del área bajo la curva. En A se observa las curvas ROC del modelo OPLS-DA para la data sensorial cruda con $AUC(C-bac) = 0.946759$ (curva roja), $AUC(C-frut) = 0.979167$ (curva amarilla), $AUC(C-chin) = 0.829861$ (curva verde). En B se observa las curvas ROC de la data sensorial optimizada con $AUC(C-bac) = 1$, $AUC(C-frut) = 1$, $AUC(C-chin) = 0.989583$.

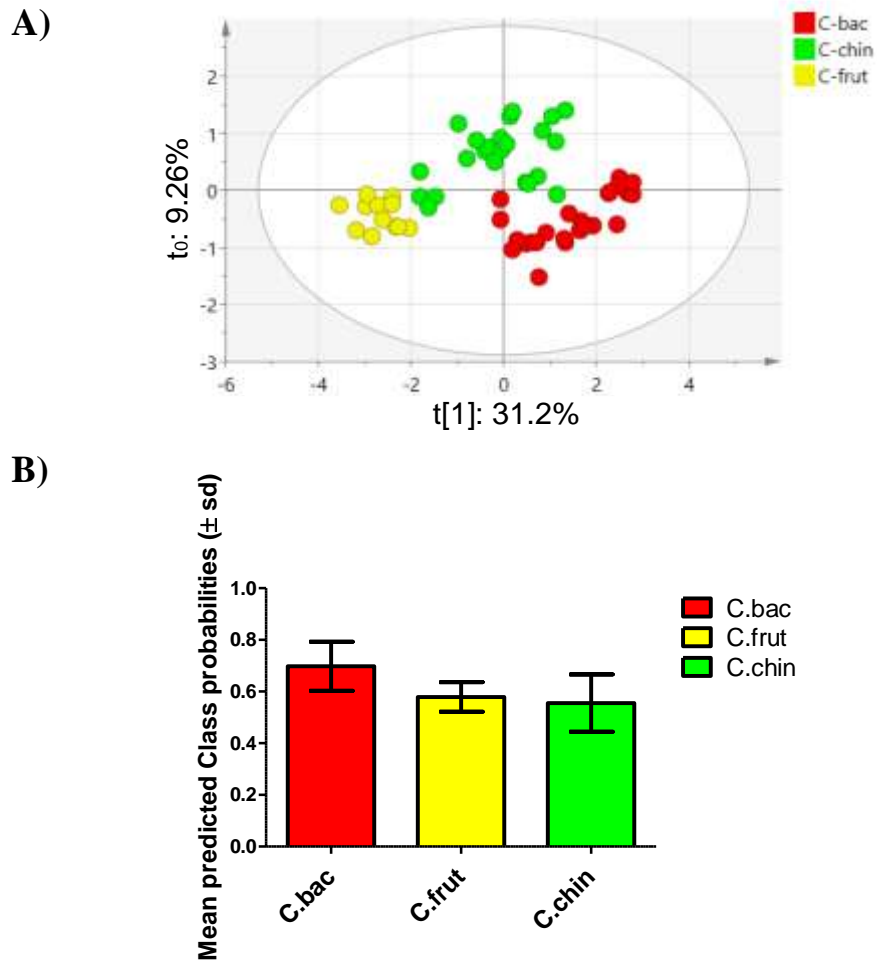


Figura 45. Estimación de la confiabilidad del modelo sensorial por OPLS-DA optimizado. En A se observa el gráfico de scores predichos a partir de la data sensorial cruda utilizando el modelo OPLS-DA de validación interna (Workset). En B se observa el diagrama de barras de las probabilidades del promedio de clases predecidas $\pm SD$, (SD = desviación estándar). Un valor superior a 0.5 indica un alto nivel de confianza de la clase predicha.

Tabla 17. Análisis (CV-ANOVA) de la data sensorial cruda y optimizada

CV-ANOVA data sensorial optimizada						
M1	SS	DF	MS	F	<i>p</i>	SD
Total corr.	118	118	1			1
Regression	86.311	8	10.7889	37.4507	4.86E-28	3.28464
Residual	31.689	110	0.288082			0.536733
CV-ANOVA data sensorial cruda						
M2	SS	DF	MS	F	<i>p</i>	SD
Total corr.	118	118	1			1
Regression	48.168	12	4.014	6.09296	5.02E-08	2.0035
Residual	69.832	106	0.658793			0.811661

7.3.7. AFM de la data sensorial optimizada

Con la data sensorial optimizada se realizó un Análisis Factorial Múltiple (AFM), analizando por cada localidad de cosecha las especies y accesiones. Los resultados muestran tendencias similares a los resultados metabolómicos observados en la data cromatográfica (PCA y OPLS-DA, ver sección 15.5.1), ya que discrimina las accesiones y especies de la misma manera que el AFM.

En la figura 46 se observa los resultados del análisis factorial múltiple (AFM) de la data sensorial original (data sensorial de los 5 ajíes de las 3 localidades). En la figura 46A se observa la distribución de las muestras (individuos) por especies y accesiones, mientras que la figura 46B muestra el círculo de correlaciones de las variables cuantitativas (descriptores del sabor y aroma). La varianza explicada del modelo es del 33.9%.

En la figura 47 muestra los resultados del análisis factorial múltiple (AFM) de la data sensorial optimizada (5 ajíes de las 3 localidades) con N = 300 datos generados por ordenador, a partir de la media y desviación estándar de los datos originales. La varianza explicada es del 26%.

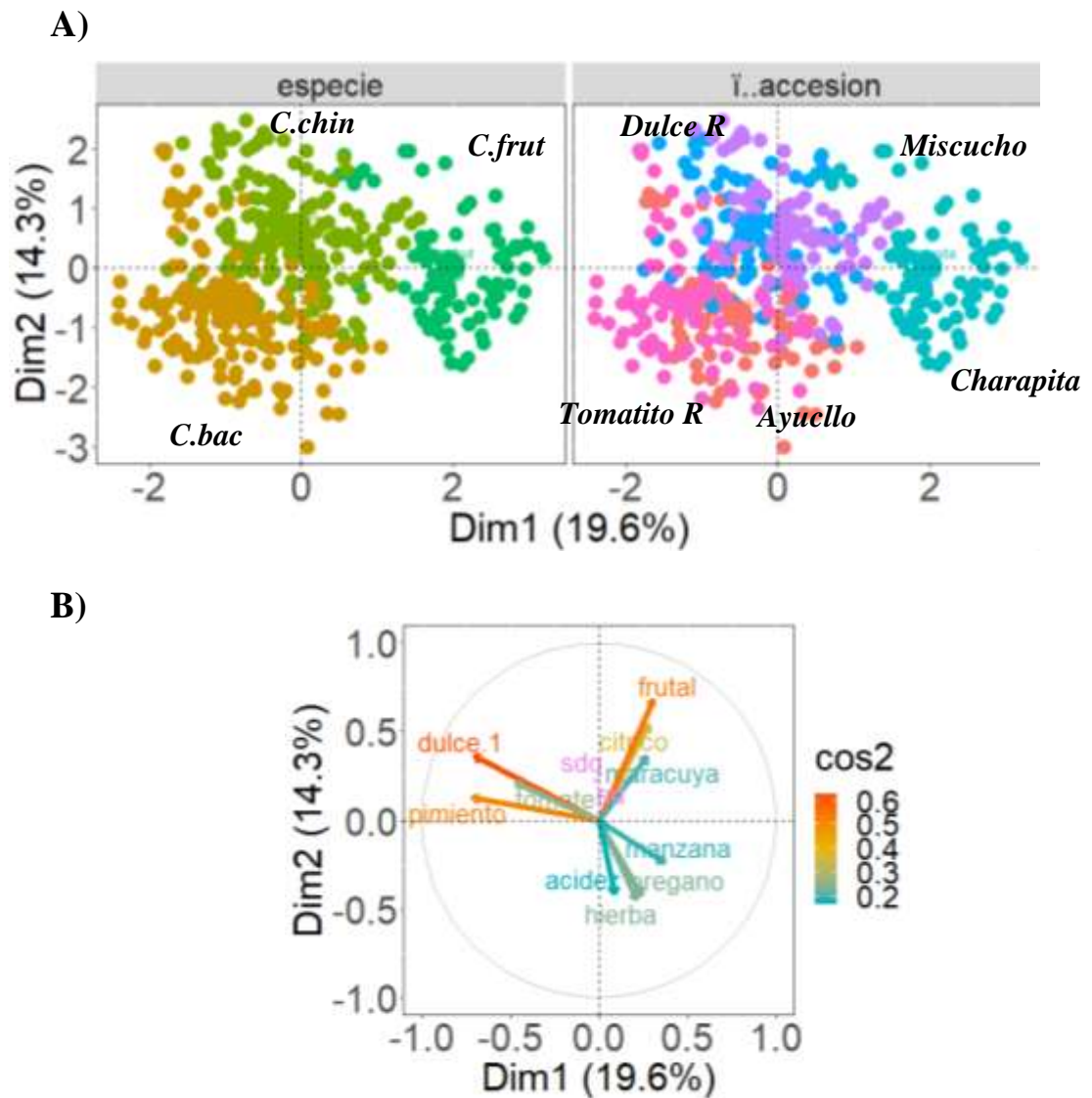
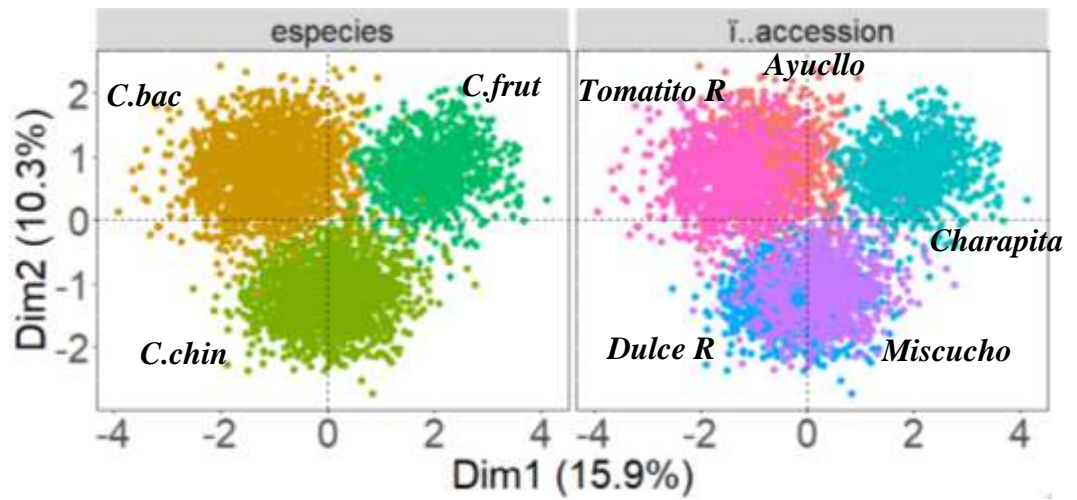


Figura 46. AFM de la data sensorial original. En A se muestra la gráfica de los individuos por especie y accesión (3 especies y 5 accesiones). En B se observa el círculo de correlación de las variables cuantitativas (descriptores del sabor y aroma). La varianza explicada es del 33.9%.

A)



B)

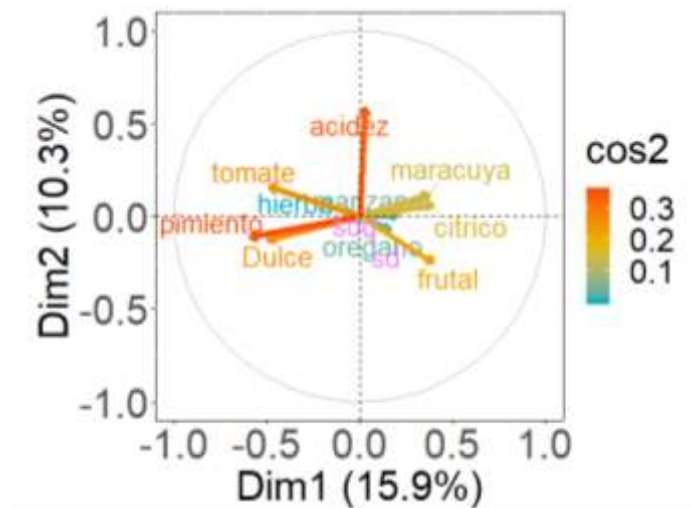


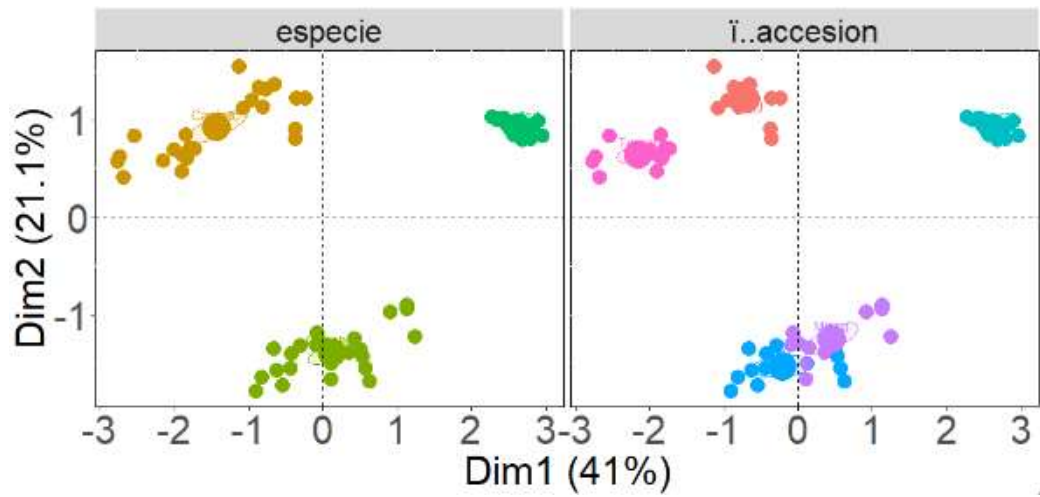
Figura 47. AFM de la data sensorial optimizada con (N = 300) datos generados por ordenador. En A se muestra la gráfica de los individuos por especie y accesión (3 especies y 5 accesiones). En B se observa el círculo de correlación de las variables cuantitativas (descriptores del sabor y aroma). La varianza explicada es del 26.3%.

En la figura 47A (izquierda superior) se observa la gráfica de individuos o nube de puntos de datos de la especie *C. baccatum* (puntos de color marrón claro). Los puntos de color verde son de la especie *C. frutescens* y los puntos en color verde claro, de la especie *C. chinense*. En la imagen de la derecha, la especie *C. baccatum* posee dos accesiones, Tomatito rojo y Ayucillo (color rosado y rojo), la especie *C. chinense* también tiene dos accesiones Miscucho y Dulce rojo (colores azul y violeta). La escala en gradiente de colores del círculo de correlaciones (Figura 47B) indica el valor del Cos^2 (coseno cuadrado o cuadrado de la coordenada) que representa la calidad de la representación de la variable sobre el mapa factorial. Un alto Cos^2 indica una buena representación de la variable sobre los componentes principales. En este caso, la variable se coloca cerca de la circunferencia del círculo de correlación. Por el contrario, un bajo Cos^2 indica que la variable no representa bien las componentes principales. En este caso la variable está cerca al centro del círculo de correlación. Los descriptores pimiento, tomate, dulce, frutal y acidez ofrecen una fuerte contribución a las componentes principales, mientras que los descriptores maracuyá y cítrico ofrecen una contribución intermedia y los descriptores hierba, manzana y orégano tienen una baja contribución.

En la figura 48 se observa la evaluación del AFM del promedio de los datos. Para realizar la gráfica, se utilizó la data generada por ordenador y se promediaron en 4 puntos por individuo, cada punto es el promedio de 75 valores ($75 \times 4 = 300$ puntos) por cada accesión. En la gráfica se observa la clasificación por especies y

acciones. La varianza explicada es del 62.3%, que es mayor que la observada de la figura 47A.

A)



B)

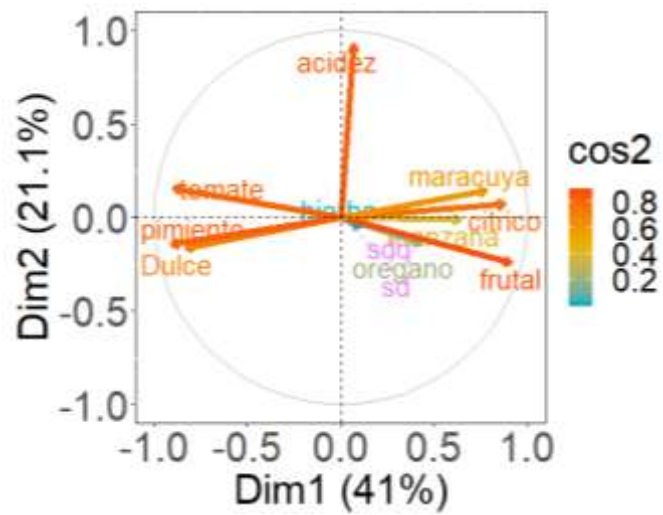


Figura 48. AFM de la data sensorial. Los puntos de la gráfica representan los valores promedios (4 puntos por acción). La varianza explicada es del 62.1%.

En la figura 49 se observa los resultados del AFM de la data sensorial optimizada de Chincha. La varianza explicada es del 70.3%. Las tablas 18 y 19 resumen los resultados de las coordenadas y calidad de las variables sobre el mapa factorial. Del círculo de correlación (Figura 49B) se observa que en el individuo *C. baccatum*, la accesión Tomatito rojo se correlaciona con los descriptores tomate ($r = -0.932$, $\text{Cos}^2 = 0.868$), pimienta ($r = -0.962$, $\text{Cos}^2 = 0.926$) y dulce ($r = -0.947$, $\text{Cos}^2 = 0.897$), que a su vez se encuentran estrechamente relacionados con Dim1 (-) y el descriptor acidez ($r = 0.878$, $\text{Cos}^2 = 0.771$), se correlaciona con la accesión Ayucello estrechamente relacionada con la Dim2 (+) (ver tablas 18 y 19).

El individuo *C. frutescens* se correlaciona con la Dim1 (+), paralelamente se correlaciona con los descriptores cítrico ($r = 0.907$, $\text{Cos}^2 = 0.822$), maracuyá ($r = 0.793$, $\text{Cos}^2 = 0.628$), orégano ($\text{Corr} = 0.758$, $\text{Cos}^2 = 0.571$), manzana ($r = 0.632$, $\text{Cos}^2 = 0.399$), hierba ($r = 0.466$, $\text{Cos}^2 = 0.217$) y frutal ($r = 0.867$, $\text{Cos}^2 = 0.751$).

El individuo *C. chinense* se encuentra fuertemente correlacionado a la Dim 2(-), se caracteriza por sus notas de orégano ($r = -0.150$, $\text{Cos}^2 = 0.023$), manzana ($r = -0.111$, $\text{Cos}^2 = 0.0123$), hierba ($r = -0.297$, $\text{Cos}^2 = 0.088$), frutal ($r = -0.459$, $\text{Cos}^2 = 0.211$), dulce ($r = -0.239$, $\text{Cos}^2 = 0.057$) y pimienta ($r = -0.168$, $\text{Cos}^2 = 0.0278$) con una baja contribución a la Dim2 (-).

En las figuras 50 y 51 se observan los resultados del AFM para la data sensorial optimizada de los ajíes de Chiclayo y Huaral, respectivamente, y muestran una

tendencia similar al perfil sensorial de Chincha (misma discriminación de accesiones y especies).

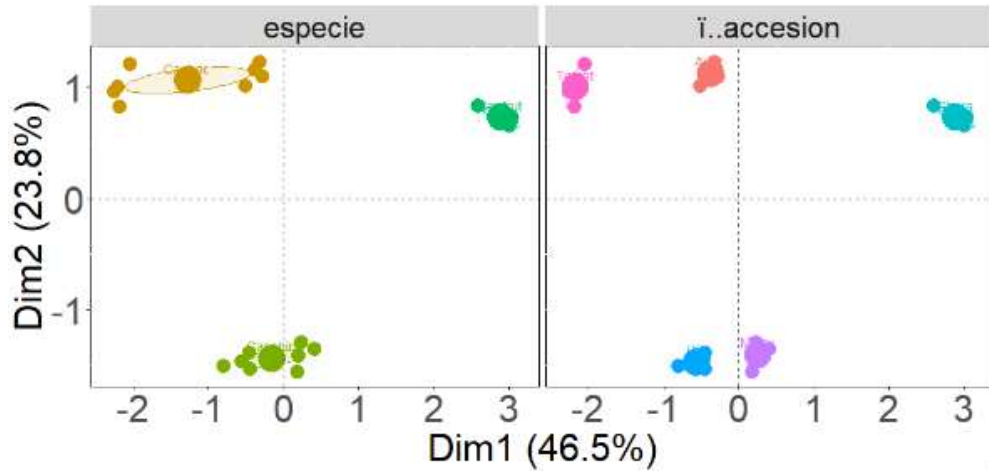
Tabla 18. Coordenadas de las variables sensoriales para los ajés de Chincha

Coordenadas	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.7557	-0.1499	0.4522	-0.0442	-0.3786
Hierba	0.4660	-0.2969	0.6952	-0.1769	0.2826
Manzana	0.6315	-0.1111	0.6708	0.1474	-0.1877
Cítrico	0.9069	0.1299	0.1683	0.0511	0.0739
Frutal	0.8668	-0.4588	0.0172	-0.0020	-0.0011
Maracuya	0.7926	0.0596	-0.2465	0.3273	0.3562
Acidez	0.3226	0.8782	0.3113	0.0116	0.0533
Dulce	-0.9473	-0.2387	0.0202	0.0101	-0.0010
Pimiento	-0.9623	-0.1667	0.1882	0.0676	-0.0101
Tomate	-0.9316	0.2579	0.2041	-0.0481	0.0222

Tabla 19. Calidad del mapa factorial (Cos2) de las variables sensoriales de los ajés de Chincha

Cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.5711	0.0225	0.2044	0.0020	0.1433
Hierba	0.2172	0.0882	0.4833	0.0313	0.0799
Manzana	0.3988	0.0123	0.4499	0.0217	0.0352
Cítrico	0.8224	0.0169	0.0283	0.0026	0.0055
Frutal	0.7513	0.2105	0.0003	0.0000	0.0000
Maracuya	0.6283	0.0036	0.0608	0.1071	0.1269
Acidez	0.1041	0.7713	0.0969	0.0001	0.0028
Dulce	0.8974	0.0570	0.0004	0.0001	0.0000
Pimiento	0.9259	0.0278	0.0354	0.0046	0.0001
Tomate	0.8678	0.0665	0.0417	0.0023	0.0005

A)



B)

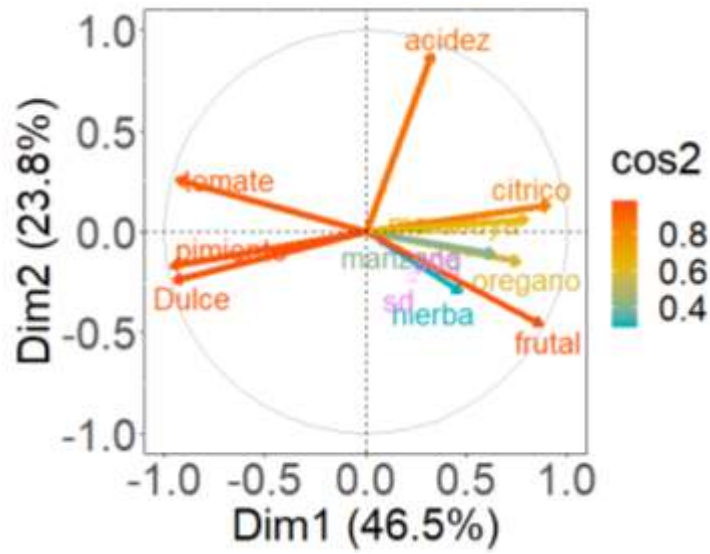
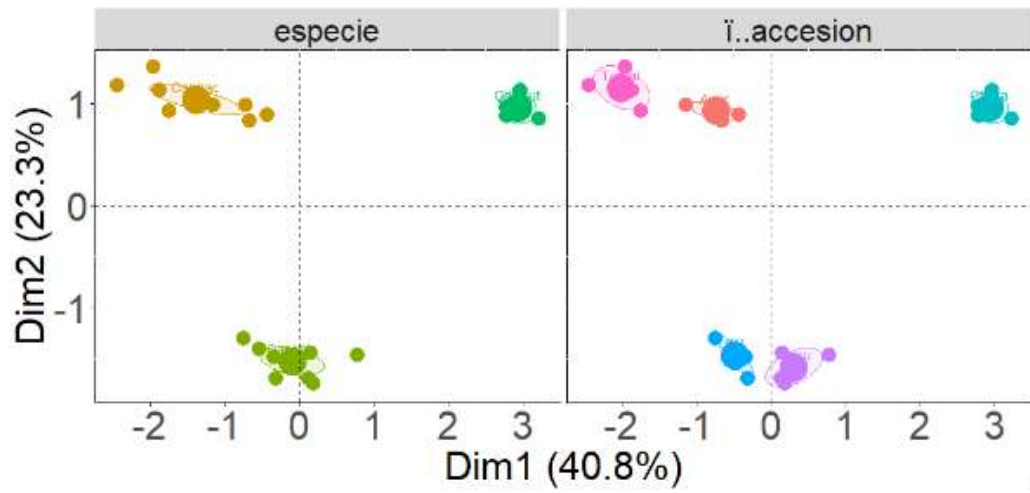


Figura 49. AFM de la data sensorial de los ajíes de Chincha. La varianza explicada es del 70.3%.

A)



B)

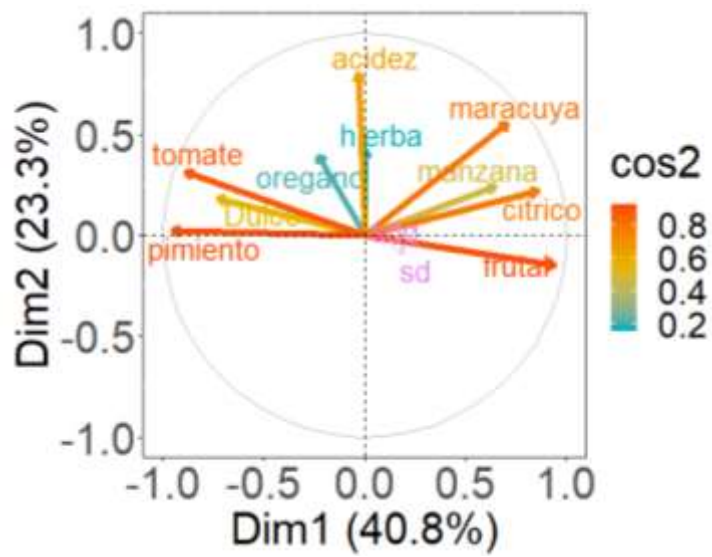
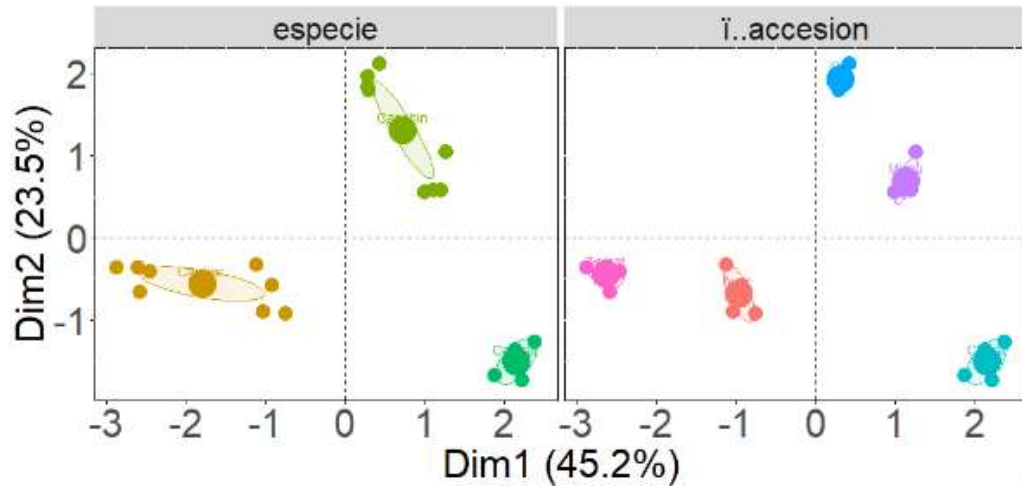


Figura 50. AFM de la data sensorial de los ajés de Chiclayo. La varianza explicada es del 64.1%.

A)



B)

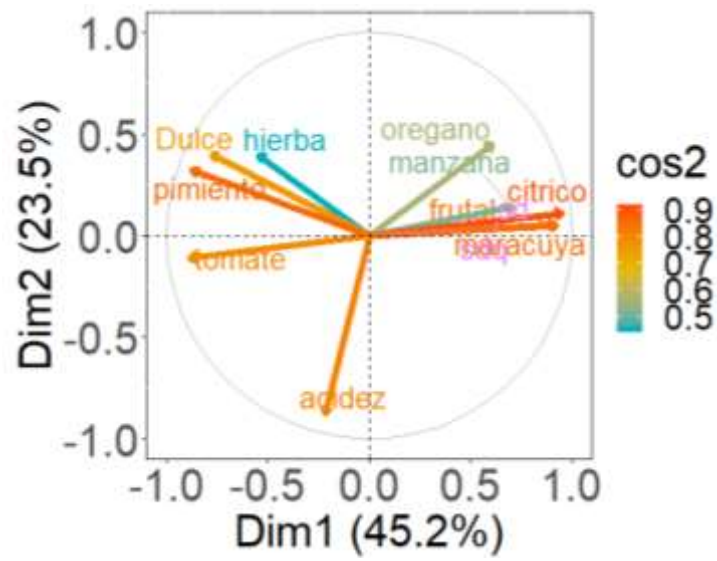


Figura 51. AFM de la data sensorial de los ajíes de Huaral. La varianza explicada es del 68.7%.

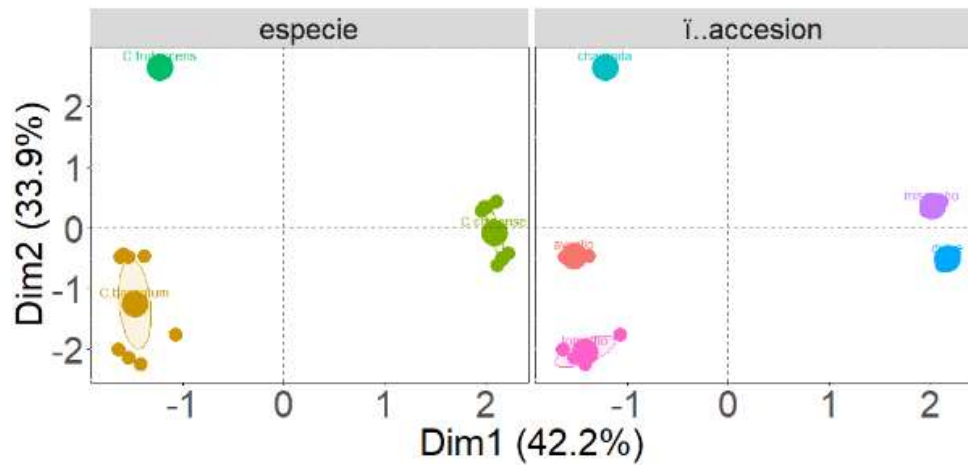
7.3.8. Correlación de las variables sensoriales y fisicoquímicas

La relación entre la información fisicoquímica de los ajíes (análisis proximal e instrumental) con los atributos sensoriales fue determinada por AFM de la data obtenida del análisis proximal e instrumental para los ajíes de Chincha (Tabla 15) y luego la data consolidada (data sensorial de Chincha + data de análisis proximal +data instrumental) (Figuras 52 y 53).

En la figura 52 se observa el resultado del AFM de las variables proximales e instrumentales realizado con los datos de la tabla 14. Se puede observar la clasificación de los individuos por accesiones y especies. Al analizar por especies se observa que el individuo *C. baccatum* se caracteriza por su contenido en grasa (betacaroteno y ácidos grasos), ($r = -0.639$, $\text{Cos}^2 = 0.408$).

El individuo *C. frutescens* se caracteriza por su contenido de proteínas ($r = 0.865$, $\text{Cos}^2 = 0.749$) estrechamente correlacionado con la variable instrumental pH ($r = 0.639$, $\text{Cos}^2 = 0.408$). El individuo *C. chinense* se encuentra estrechamente relacionado con el contenido de humedad ($r = 0.944$, $\text{Cos}^2 = 0.891$) y grados Brix ($r = 0.783$, $\text{Cos}^2 = 0.613$). En la figura 53 se observa el resultado del AFM de las dos tablas de datos consolidados (variables sensoriales y fisicoquímicas) (Tablas 20 y 21).

A)



B)

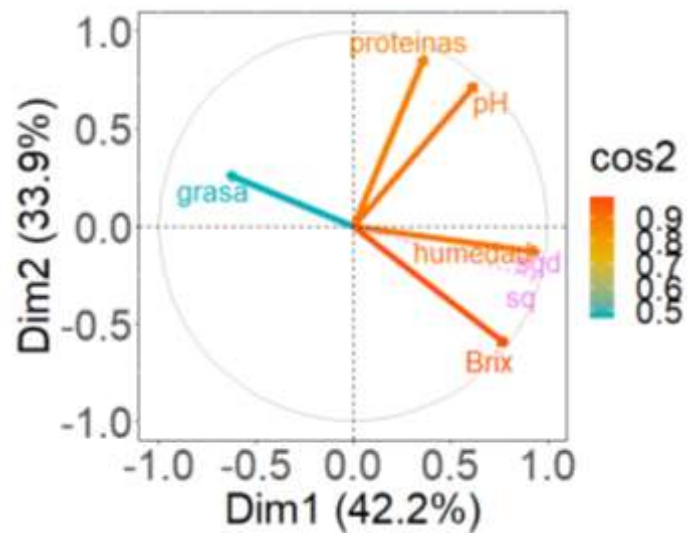


Figura 52. AFM, variables fisicoquímicas de los ajjes de Chincha.

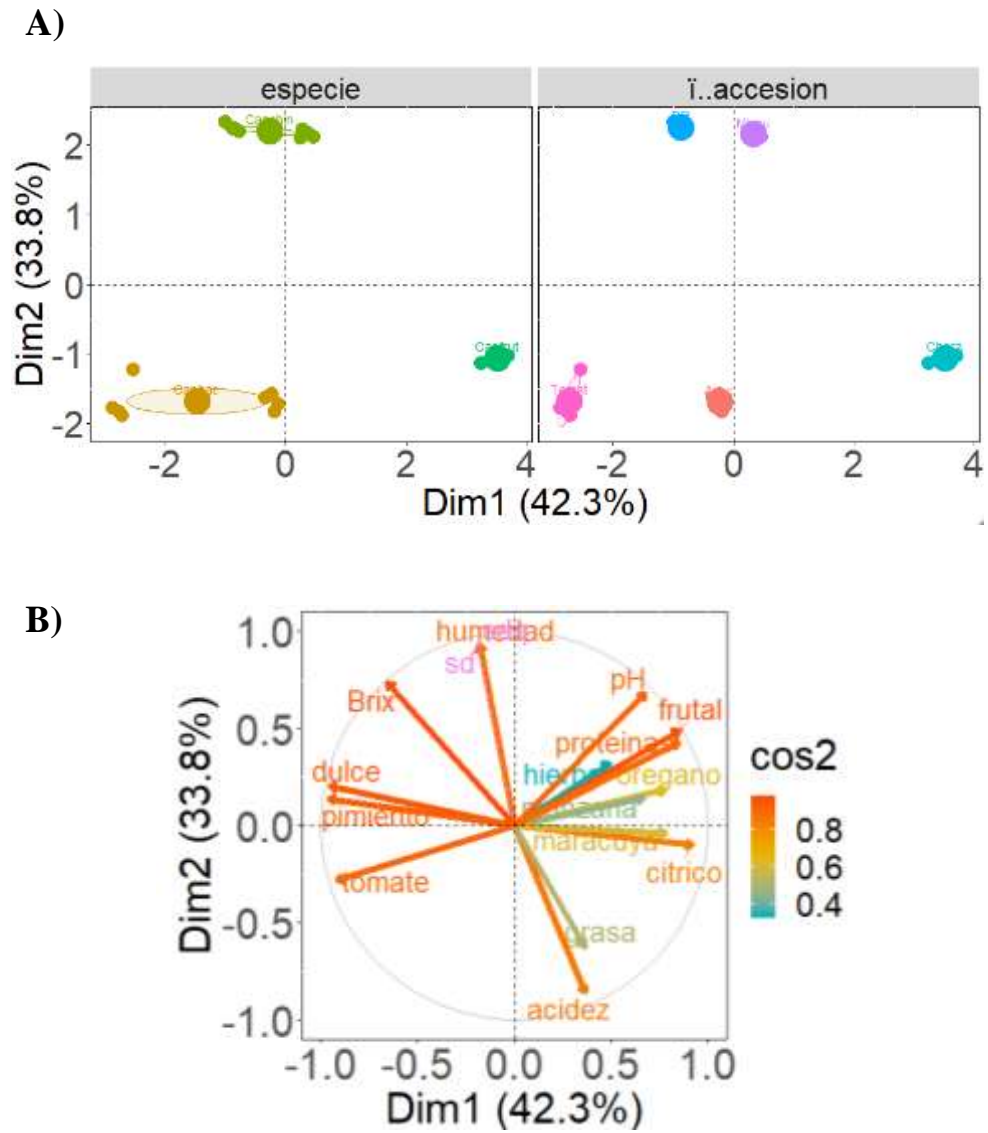


Figura 53. AFM, correlación de la data sensorial y fisicoquímica de los ajés de Chíncha.

Tabla 20. Coordenadas de las variables químicas e instrumental para los ajíes de Chincha

Coordenadas	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Humedad	0.9440	-0.1304	0.2659	0.0178	-0.0402
Grasa	-0.6393	0.2661	0.7207	-0.0310	0.0084
Proteínas	0.3651	0.8656	0.1716	0.2847	-0.0687
pH	0.6266	0.7255	0.0185	-0.2537	0.1277
Brix	0.7831	-0.6055	0.1033	0.0881	0.0123

Tabla 21. Calidad del mapa factorial (Cos2) de las variables químicas e instrumental para los ajíes de Chincha

Cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Humedad	0.8912	0.0170	0.0707	0.0003	0.0016
Grasa	0.4087	0.0708	0.5194	0.0010	0.0001
Proteínas	0.1333	0.7493	0.0294	0.0811	0.0047
pH	0.3926	0.5264	0.0003	0.0644	0.0163
Brix	0.6132	0.3666	0.0107	0.0078	0.0002

En la figura 53 se observa el AFM de la data consolidada (data sensorial, químicas e instrumentales) para los ajíes de Chincha. Las tablas 22 y 23 resumen los resultados de las coordenadas y calidad de la representación de las variables sobre el círculo de correlación (Figura 53B). Del círculo de correlaciones (Figura 53B), las variables sensoriales del olor, proteínas y pH están fuertemente correlacionados con la accesión Charapita en el gráfico de individuos. También, las variables del sabor (pimiento, dulce y tomate) se correlacionan con los grados Brix y con la accesión Tomatito rojo en el gráfico de individuos, las grasas se correlacionan con la variable acidez que simultáneamente se correlacionan con la accesión Ayuello en el gráfico de individuos.

Tabla 22. Coordenadas de las variables químicas, instrumental y sensoriales para los ajíes de Chincha

Coordenadas	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.7734	0.1840	0.4017	-0.0343	-0.3565
Hierba	0.4859	0.3198	0.6646	-0.1682	0.1616
Manzana	0.6612	0.1428	0.6225	0.1535	-0.1618
Cítrico	0.9147	-0.0996	0.1194	0.0250	0.0608
Frutal	0.8494	0.4850	-0.0294	-0.0285	0.0298
Maracuya	0.7778	-0.0384	-0.2778	0.2615	0.4471
Acidez	0.3638	-0.8550	0.3145	0.0004	0.0738
Dulce	-0.9499	0.2019	0.0570	0.0397	-0.0341
Pimiento	-0.9538	0.1389	0.2306	0.1008	-0.0247
Tomate	-0.9142	-0.2795	0.2549	-0.0163	-0.0022
Humedad	-0.1797	0.9280	0.2778	-0.0014	0.0577
Grasa	0.3610	-0.6215	0.6899	0.0584	0.0180
Proteínas	0.8461	0.4281	0.1227	-0.2774	0.0504
pH	0.6719	0.6796	-0.0281	0.2475	-0.0996
Brix	-0.6528	0.7349	0.1566	-0.0786	0.0284

Tabla 23. Calidad del mapa factorial (Cos2) de las variables químicas, instrumental y sensoriales para los ajíes de Chincha

Cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.5982	0.0338	0.1614	0.0012	0.1271
Hierba	0.2361	0.1023	0.4417	0.0283	0.0261
Manzana	0.4371	0.0204	0.3875	0.0236	0.0262
Cítrico	0.8367	0.0099	0.0143	0.0006	0.0037
Frutal	0.7215	0.2353	0.0009	0.0008	0.0009
Maracuya	0.6049	0.0015	0.0772	0.0684	0.1999
Acidez	0.1323	0.7310	0.0989	0.0000	0.0054
Dulce	0.9023	0.0408	0.0033	0.0016	0.0012
Pimiento	0.9098	0.0193	0.0532	0.0102	0.0006
Tomate	0.8357	0.0781	0.0650	0.0003	0.0000
Humedad	0.0323	0.8611	0.0771	0.0000	0.0033
Grasa	0.1303	0.3863	0.4760	0.0034	0.0003
Proteínas	0.7159	0.1833	0.0150	0.0770	0.0025
pH	0.4514	0.4618	0.0008	0.0613	0.0099
Brix	0.4261	0.5401	0.0245	0.0062	0.0008

7.3.9. Normalización de la data sensorial optimizada de Chincha

La normalización de la data sensorial fue realizada utilizando el método de normalización EigenMS para mejorar la clasificación de las variables por especies y accesiones. Para realizar la normalización se preparó la tabla de datos sensoriales colocando las variables cuantitativas sensoriales (10 descriptores) en columnas y las observaciones en filas (4 observaciones por accesión). Para la normalización se utilizó la plataforma del software online NOREVA que tiene a disposición 20 métodos de normalización. El procedimiento de normalización consta de cuatro etapas: **1)** en esta etapa la tabla de datos en formato *.CSV se sube a la nube señalando el tipo de data de trabajo con muestras con QC (control de calidad) o sin QC. Para este análisis se eligió la 2da opción (muestras sin QC) **2)** la segunda etapa es la etapa de preprocesamiento de la data en la que se mantuvo los parámetros por defecto del software NOREVA ((1) Número de vecinos = 10, (2) Número máximo de datos que faltan permitidos en cualquier fila (%) = 50% , (3) Número máximo de datos faltantes permitidos en cualquier columna (%) = 80%, (4) El mayor bloque de metabolitos imputados = 1500, **3)** En la tercera etapa, se eligió el método de normalización EigenMS, los resultados fueron descargados en formato *.CSV, **4)** el desempeño se evaluó por la capacidad del método para reducir la variación intragrupo entre muestras. Para ello NOREVA evalúa el PCA antes y después de la normalización. En la figura 54 se observa el diagrama boxplots de la data sensorial de la categoría especies antes y después de la normalización. El diagrama de boxplots es una visualización gráfica que describe simultáneamente varias

características importantes de un conjunto de datos, como el centro, la dispersión, la desviación de la simetría y la identificación de observaciones inusuales o valores atípicos. Un diagrama de boxplots muestra los tres cuartiles, el mínimo y el máximo de los datos en una caja rectangular, alineados horizontal o verticalmente (115,116). La figura 55, muestra los PCA antes y después de la normalización observándose un buen desempeño del método para clasificar las muestras por especies. En la figura 56 se observan los diagramas boxplots de la data sensorial de la categoría por accesiones antes y después de la normalización. La figura 57 muestra los PCA antes y después de la normalización, observándose un buen desempeño del método para clasificar las muestras por accesiones.

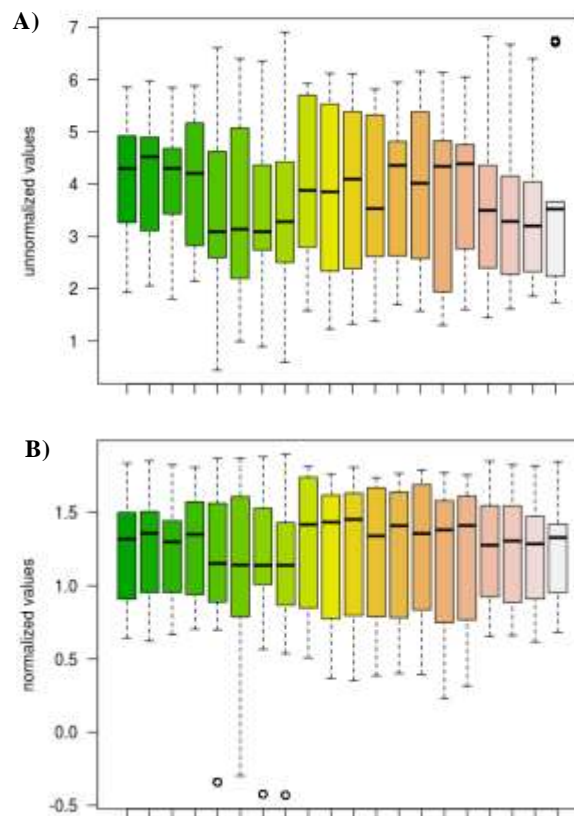


Figura 54. Normalización de la data sensorial por especies utilizando el método de normalización EigenMS. En A y en B se observa el diagrama boxplots de 20 observaciones para las tres especies antes y después de la normalización respectivamente.

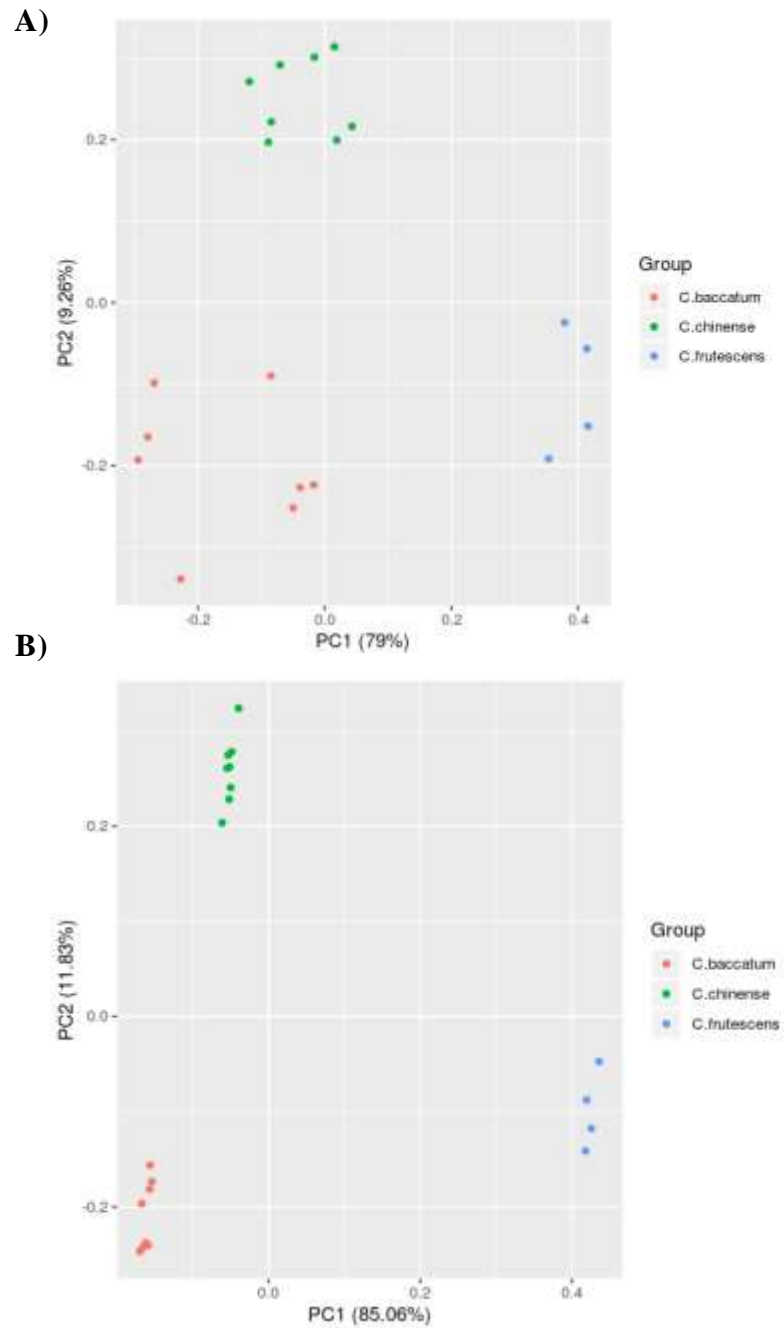


Figura 55. Evaluación de la capacidad del método EigenMS para reducir la variación intragrupo entre especies, se muestra el PCA antes y después de la normalización con una mejorada clasificación, la varianza explicada aumenta del 88.26 a 96.89%.

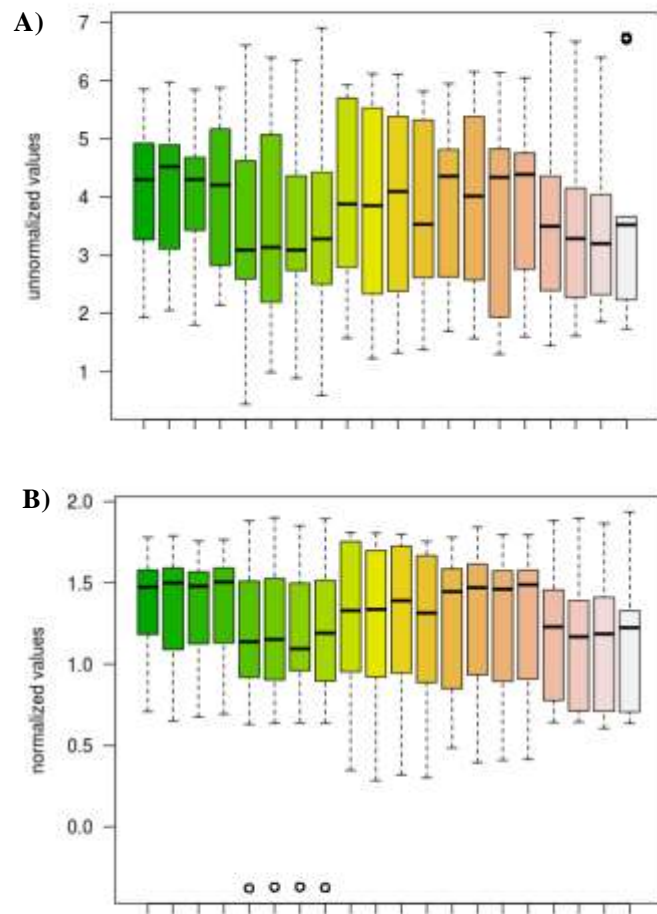


Figura 56. Normalización de la data sensorial por accesiones utilizando el método de normalización EigenMS. En A y en B se observa el diagrama boxplots de 20 observaciones para las 5 accesiones antes y después de la normalización respectivamente.

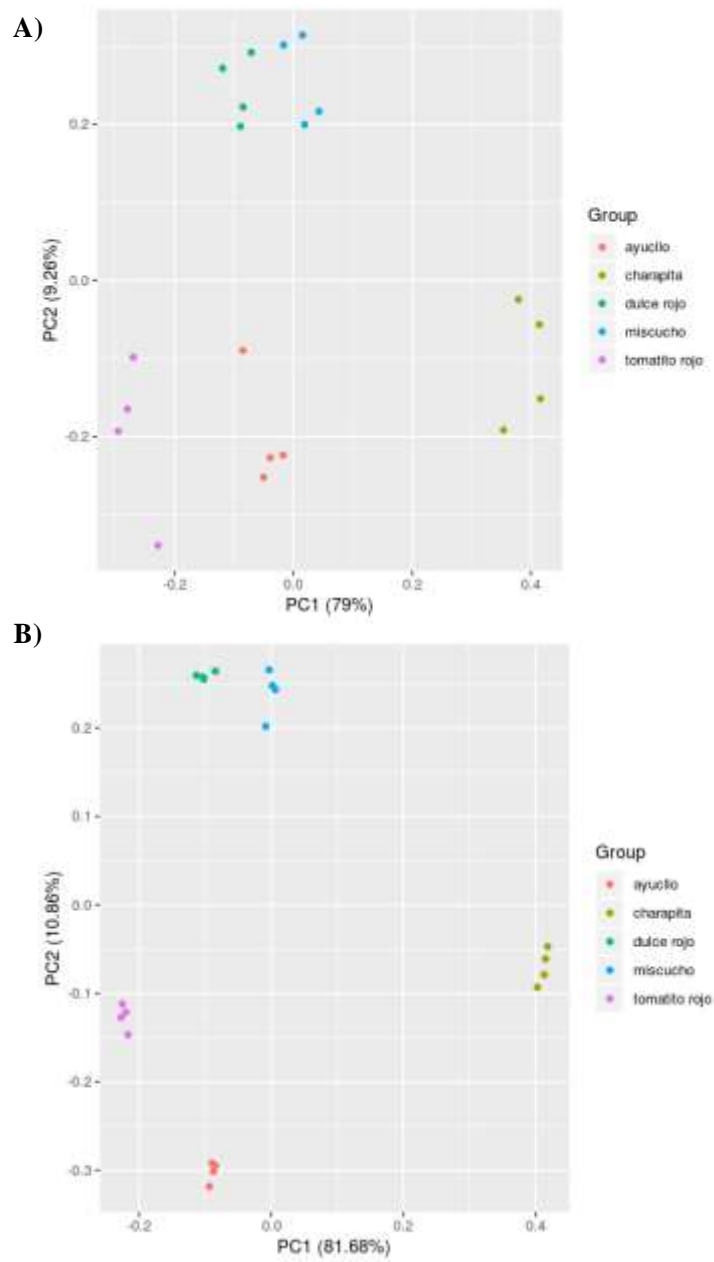


Figura 57. Evaluación de la capacidad del método EigenMS para reducir la variación intragrupo entre accesiones, se muestra el PCA antes y después de la normalización, la varianza explicada aumenta del 88.26 a 92.54%.

En la figura 58 se observa los gráficos ROC en el que se evaluó el área bajo la curva como medida de la precisión en la clasificación categórica. Un área bajo la curva $AUC > 0.7$ es considerada como una buena clasificación entre las clases. Cuando se normalizó por accesiones, se observó en el gráfico ROC (Figura 58B), un $AUC = 1$, por lo que se considera que la clasificación por accesiones ofrece un mejor performance en el método de Normalización EigenMS que por especies.

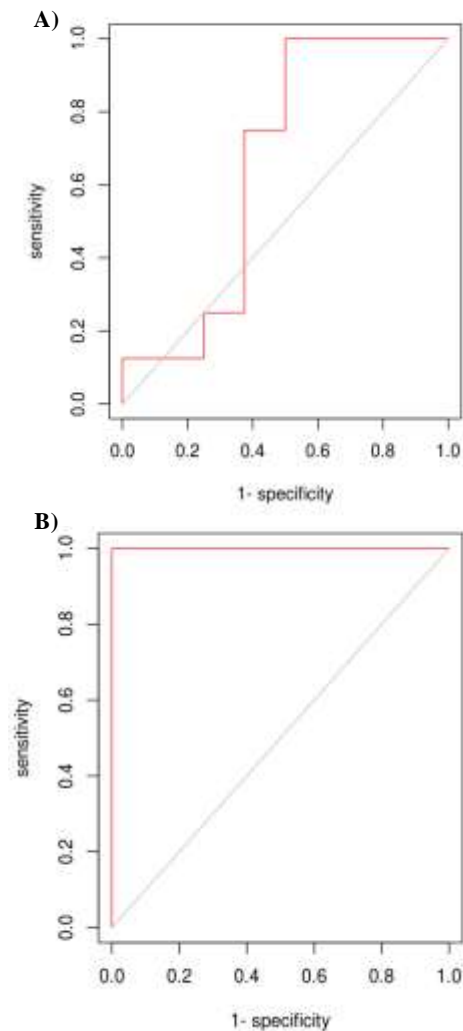


Figura 58. Gráficos ROC para la evaluación de la influencia del método de normalización EigenMS en la precisión de la clasificación. Para la clasificación por especies se observa un Área bajo la curva (AUC): $AUC = 0.6502$, y por accesiones un $AUC = 1$. Un $AUC > 0.7$ se considera como buena precisión para la clasificación.

7.3.10. AFM de la data sensorial optimizada y normalizada de Chíncha

En las figuras 59 y 60 se muestran los resultados del análisis AFM de la data sensorial optimizada de Chíncha normalizada por el método EigenMS por especies y accesiones. La data normalizada (accesiones) fue utilizada en el análisis multibloque para determinar los descriptores químicos del sabor y aroma, ya que esta data presentó un mejor performance (AUC = 1).

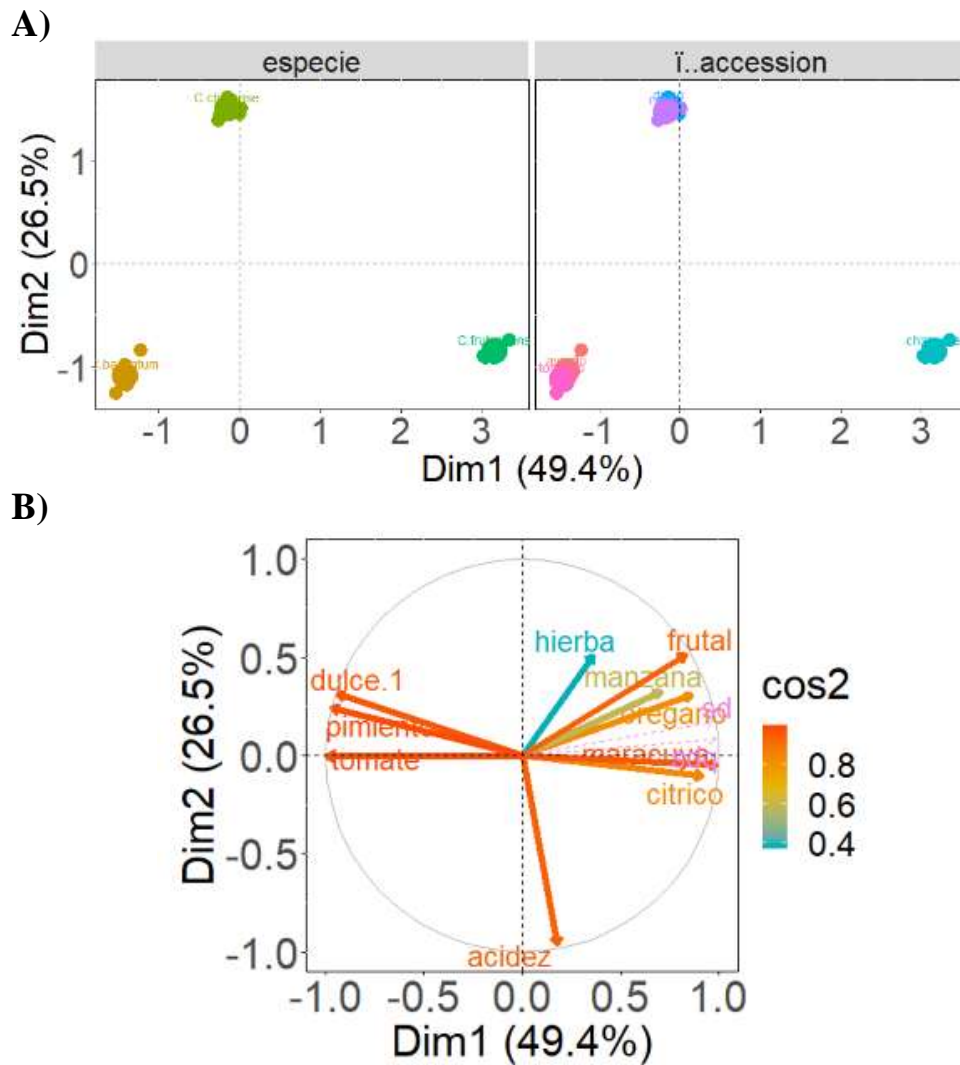


Figura 59. AFM de la data sensorial normalizada por especies por el método EigenMS de los ajés de Chíncha. La varianza explicada es del 75.9%.

En las tablas 24-26 se muestran los resultados cuantitativos de las coordenadas y calidad de representación de las variables por especies y las tablas 27-29 por accesiones.

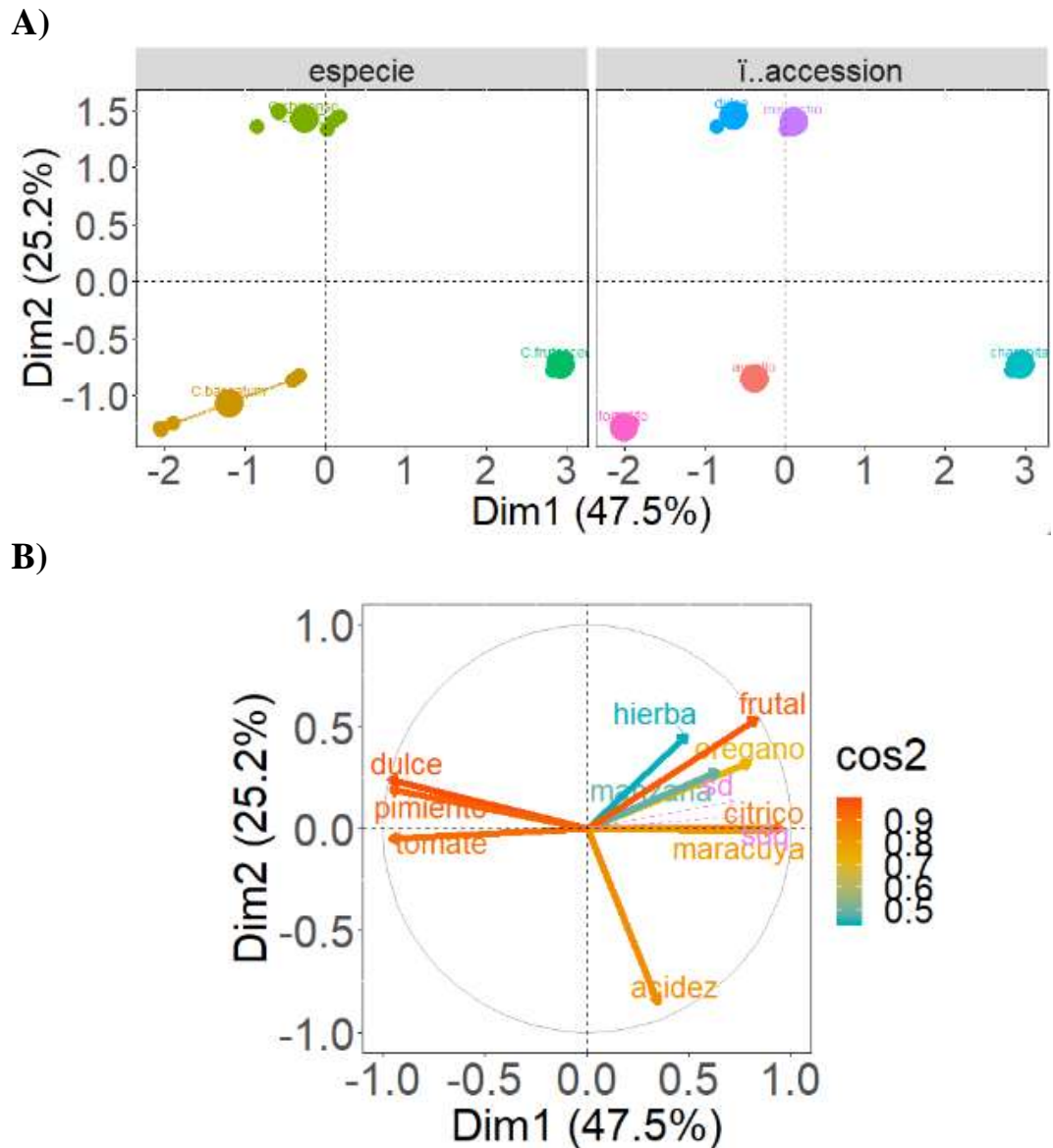


Figura 60. AFM de la data sensorial normalizada por accesiones por el método EigenMS de los ajíes de Chincha. La varianza explicada es del 72.7%.

Tabla 24. Contribuciones a las dimensiones por especies

Contribuciones	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	6.1914	1.4551	0.3920	0.0051	4.8678
Hierba	1.0638	4.0315	8.7303	0.0008	56.3211
Manzana	4.1388	1.6446	0.0184	0.0743	23.5048
Cítrico	7.0072	0.1670	0.0981	0.0591	5.6762
Frutal	5.7629	4.0885	0.0497	0.0000	0.2970
Maracuya	8.1497	0.0347	0.0072	0.0112	0.0167
Acidez	0.3618	19.6825	0.7657	0.0000	0.4116
Dulce	10.0557	2.0610	0.1441	0.0006	0.0149
Pimiento	10.7142	1.2443	0.0003	0.0001	0.7820
Tomate	11.3503	0.0006	0.0712	0.0021	0.0464

Tabla 25. Calidad del mapa factorial (Cos2) por especies

Cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.7346	0.0925	0.0089	0.0001	0.0273
Hierba	0.1262	0.2563	0.1992	0.0000	0.3158
Manzana	0.4910	0.1045	0.0004	0.0016	0.1318
Cítrico	0.8313	0.0106	0.0022	0.0012	0.0318
Frutal	0.6837	0.2599	0.0011	0.0000	0.0017
Maracuya	0.9669	0.0022	0.0002	0.0002	0.0001
Acidez	0.0315	0.9172	0.0128	0.0000	0.0017
Dulce	0.8746	0.0960	0.0024	0.0000	0.0001
Pimiento	0.9319	0.0580	0.0000	0.0000	0.0032
Tomate	0.9872	0.0000	0.0012	0.0000	0.0002

Tabla 26. Coordenadas de las variables en las dimensiones por especies

Coordenadas	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.8571	0.3041	-0.0946	0.0103	-0.1652
Hierba	0.3553	0.5062	0.4463	0.0042	0.5619
Manzana	0.7007	0.3233	-0.0205	0.0395	-0.3630
Cítrico	0.9118	-0.1030	0.0473	-0.0352	0.1784
Frutal	0.8269	0.5098	0.0337	0.0006	0.0408
Maracuya	0.9833	-0.0469	-0.0128	0.0153	-0.0097
Acidez	0.1774	-0.9577	0.1132	-0.0002	-0.0411
Dulce	-0.9352	0.3099	0.0491	0.0029	0.0078
Pimiento	-0.9654	0.2408	-0.0024	-0.0015	-0.0567
Tomate	-0.9936	-0.0052	0.0345	0.0056	0.0138

Tabla 27. Contribuciones a las dimensiones por accesiones

Contribuciones	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	5.0564	1.5696	5.4967	0.1018	29.9386
Hierba	1.8794	3.1738	11.2437	1.2554	11.1537
Manzana	3.2735	1.1759	10.2928	0.5170	45.5928
Cítrico	7.3132	0.0001	1.9000	0.2235	0.1546
Frutal	5.5194	4.4141	0.0388	0.0003	0.8870
Maracuya	6.5773	0.0012	0.5646	5.6398	7.6391
Acidez	1.4555	16.5529	5.2253	0.0360	0.1417
Dulce	11.2532	1.3445	0.2100	0.1008	0.8676
Pimiento	11.0885	0.8637	1.3373	0.2970	0.0110
Tomate	11.1841	0.0512	2.4688	0.0374	0.4983

Tabla 28. Calidad del mapa factorial (Cos2) por accesiones

Cos2	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.6248	0.1026	0.2345	0.0026	0.0326
Hierba	0.2322	0.2076	0.4796	0.0323	0.0122
Manzana	0.4045	0.0769	0.4391	0.0133	0.0497
Cítrico	0.9037	0.0000	0.0811	0.0058	0.0002
Frutal	0.6820	0.2887	0.0017	0.0000	0.0010
Maracuya	0.8127	0.0001	0.0241	0.1453	0.0083
Acidez	0.1203	0.7244	0.1492	0.0006	0.0001
Dulce	0.9305	0.0588	0.0060	0.0017	0.0006
Pimiento	0.9169	0.0378	0.0382	0.0051	0.0000
Tomate	0.9248	0.0022	0.0705	0.0006	0.0004

Tabla 29. Coordenadas de las variables en las dimensiones por accesiones

Coordenadas	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Orégano	0.7904	0.3204	0.4842	-0.0512	0.1806
Hierba	0.4819	0.4556	0.6926	-0.1798	0.1103
Manzana	0.6360	0.2773	0.6626	0.1154	-0.2229
Cítrico	0.9506	0.0027	0.2847	0.0759	-0.0130
Frutal	0.8258	0.5373	0.0407	-0.0029	0.0311
Maracuya	0.9015	-0.0088	-0.1552	0.3812	0.0912
Acidez	0.3469	-0.8511	0.3862	0.0249	0.0102
Dulce	-0.9646	0.2426	0.0774	0.0417	0.0252
Pimiento	-0.9575	0.1944	0.1954	0.0715	-0.0028
Tomate	-0.9616	-0.0473	0.2655	-0.0254	0.0191

7.3.11. Diagramas Heatmap y PLS-DA de la data sensorial de Chincha

En las figuras 61 y 62 se observa el diagrama Heatmap de la data sensorial de Chincha de las clases categóricas por accesiones y especies, respectivamente, que fue construido a partir de la data sensorial normalizada por el método EigenMS. Los diagramas fueron diseñados a partir del modelo PLS-DA de las VIP (variables de importancia) de los descriptores del sabor y aroma utilizando el software online MetaboAnalyst 3.0. La imagen coloreada comunica claras diferencias entre las especies y accesiones de los ajíes, a través del agrupamiento jerárquico (HCA) realizado tanto para filas y columnas en la matriz de datos sensorial. En las filas se observa el agrupamiento jerárquico por descriptores del sabor y aroma y en columnas el agrupamiento jerárquico de las accesiones y especies. Las observaciones han sido ordenadas de acuerdo a su similaridad. La escala de colores muestra los valores altos y bajos con valores similares uno cerca del otro. Como medida de distancia se utilizó la distancia Euclidiana y como algoritmo de agrupamiento se utilizó el algoritmo Ward. Estos diagramas son complementarios con los diagramas del AFM (Figuras 59 y 60) que permite evaluar de manera integral las correlaciones de las variables sensoriales con sus respectivos individuos. Por ejemplo, en la figura 61, la especie *C. frutescens* posee los siguientes descriptores: acidez, hierba, cítrico, frutal, manzana y maracuyá con una alta intensidad (+2) en la escala de intensidad, mientras que posee una muy baja intensidad de los descriptores pimienta, dulce y tomate (-2). La especie *C. baccatum* se caracteriza por los descriptores tomate, dulce, pimienta y acidez, con una intensidad intermedia (+1) en la escala de intensidad, mientras que posee una baja

intensidad en los otros descriptores. Estas observaciones están de acuerdo con los resultados del AFM (Figuras 59A y 59B), y juntos ofrecen una mejor interpretación del análisis sensorial. De igual manera se puede interpretar los resultados para el caso de las accesiones (Figuras 60A, 60B). En la figura 63 se observa la gráfica de scores de valores de importancia (VIP). Los valores de importancia en la Proyección (VIP) estiman el poder explicativo de una variable sobre el conjunto de variables predecidas “Y” o proyección, utilizada en un modelo PLS y a menudo se utiliza para la selección de las variables. Una variable con un puntaje VIP cercano o mayor a 1 puede considerarse importante en un modelo dado. Las variables con scores VIP significativamente menores a 1 (uno) son menos importantes (117). Las variables de importancia del modelo PLS-DA por especies son los descriptores tomate, dulce, pimienta y frutal ($VIP > 1$) y por accesiones son los descriptores pimienta, hierba, manzana, cítrico, acidez, orégano y dulce ($VIP > 1$). También se puede interpretar del diagrama de score VIP por especies, los descriptores que destacan de la especie *C. baccatum* son tomate dulce y pimienta, para *C. chinense* el descriptor destacado es hierba y para *C. frutescens* los descriptores destacados son frutal, maracuyá, cítrico, orégano, manzana y acidez, según la escala de intensidad o VIP (Figura 63). Similarmente, en el diagrama de score VIP por accesiones, la accesión Ayucllo destaca por sus notas de acidez, manzana y hierba, la accesión Tomatito destaca por sus notas de dulce, pimienta y tomate, la accesión Miscucho destaca por sus notas de manzana frutal y maracuyá (intensidad media), la accesión Dulce rojo destaca por sus notas de dulce, hierba y pimienta (intensidad media), la accesión Charapita destaca por sus notas de orégano, cítrico, frutal y

maracuyá. En las figuras 64 y 65 se observan los diagramas de scores y loadings 3D-PLS-DA, que demuestran una excelente clasificación por especies y accesiones (R^2 y $Q^2 > 0.7$).

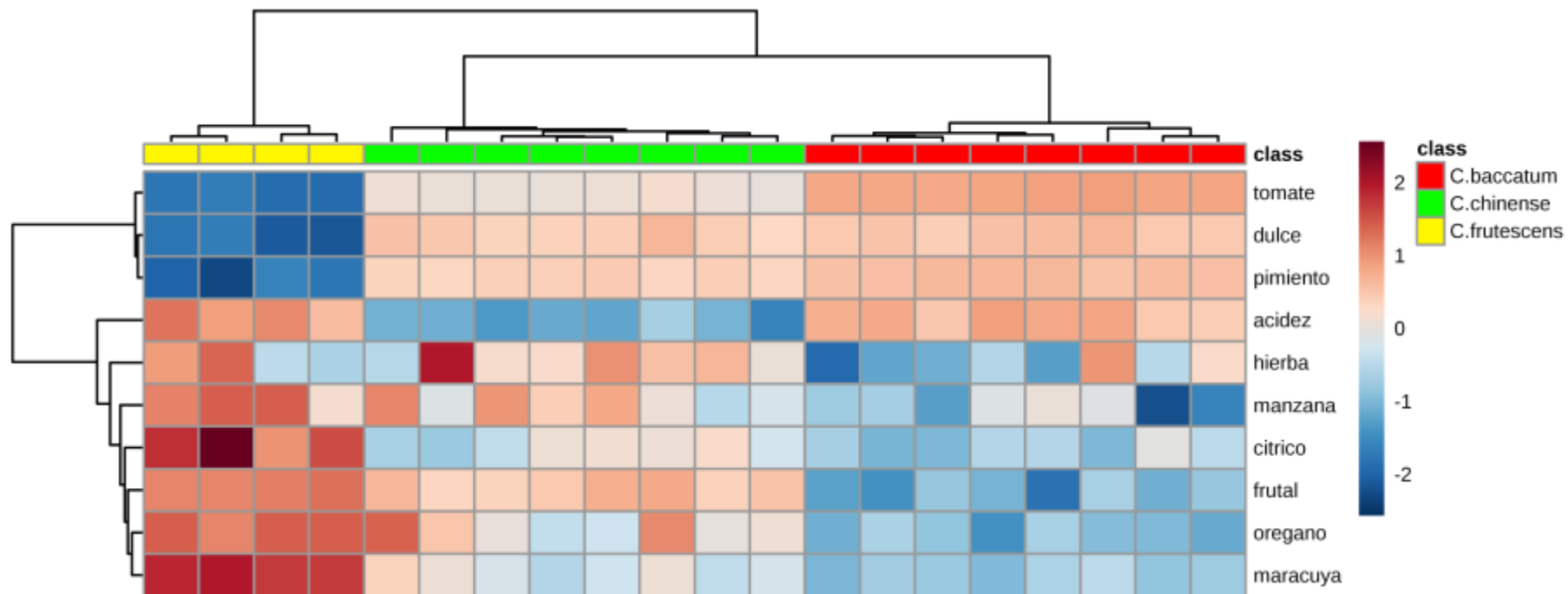


Figura 61. Diagrama Heatmap por especies

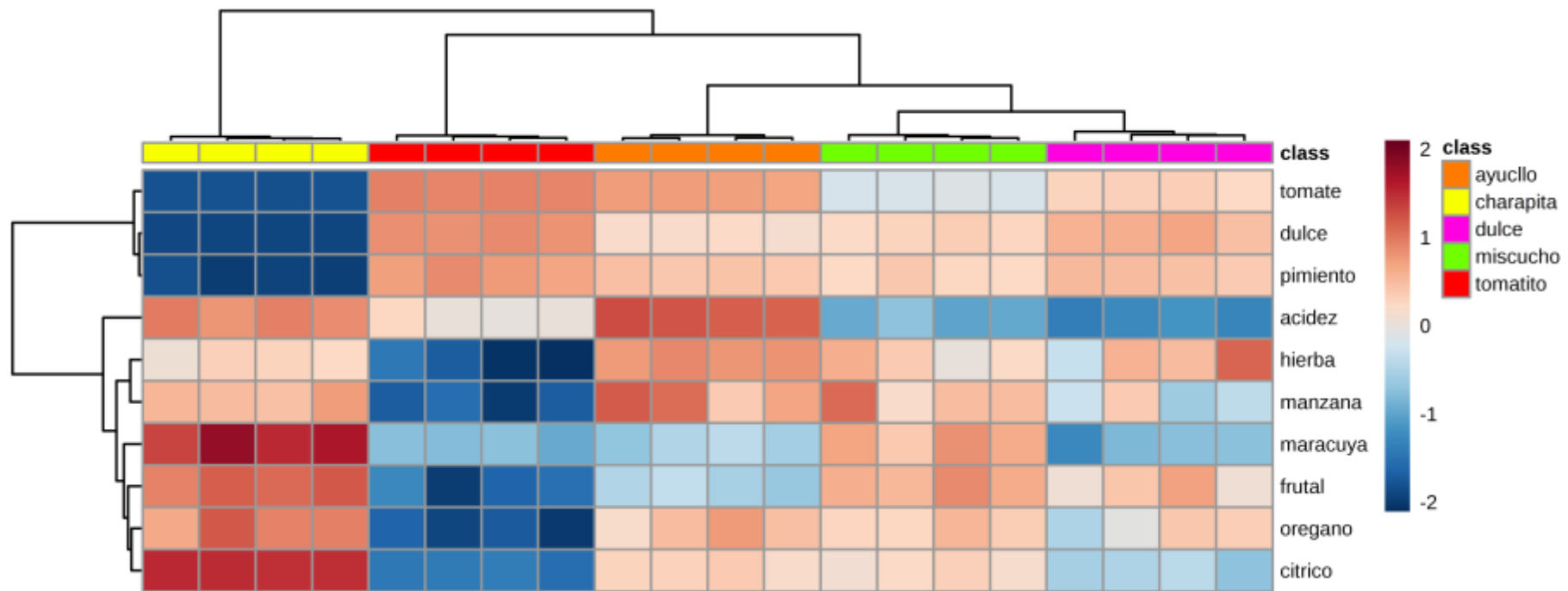


Figura 62. Diagrama Heatmap por accesiones

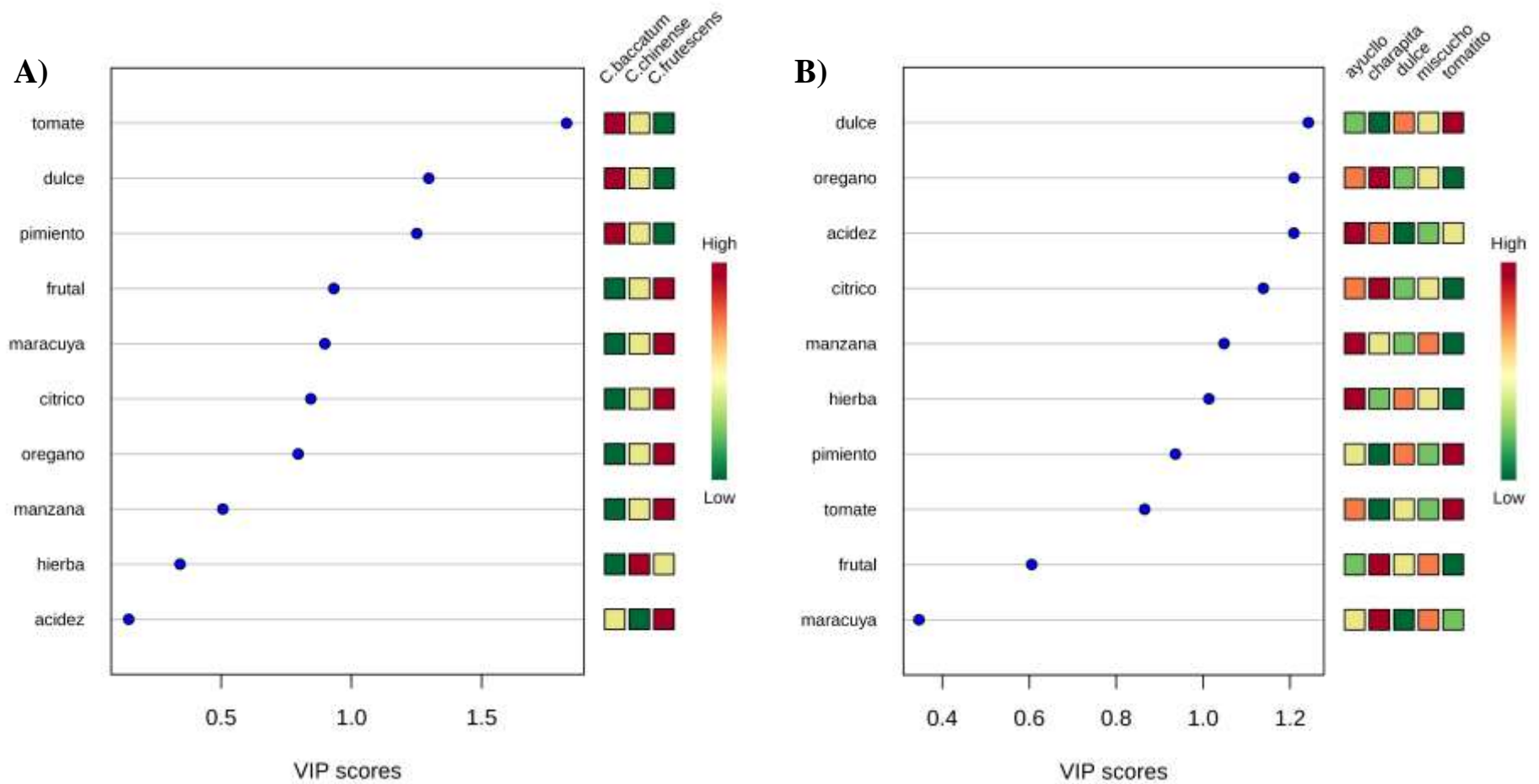


Figura 63. Principales valores de importancia (VIP) por especies (A) y accesiones (B) obtenidos a partir del modelo PLS-DA de la data sensorial de Chinchá. La gráfica VIP indica los descriptores más discriminantes en orden descendente de importancia.

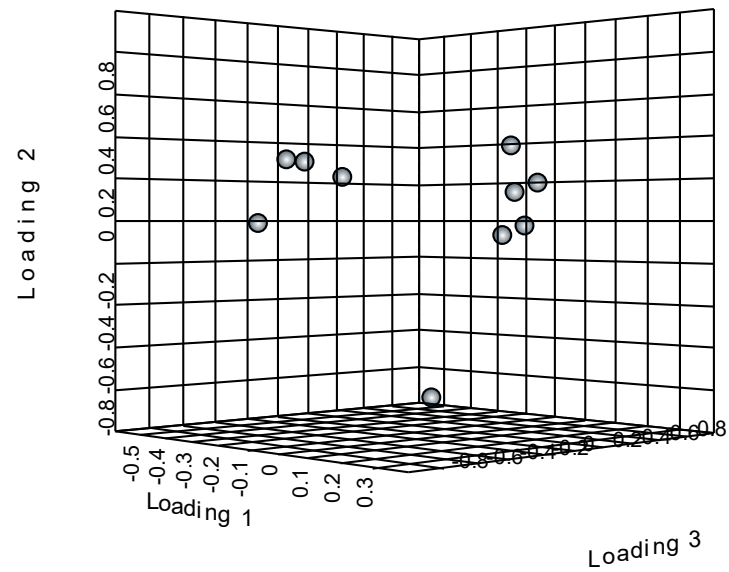
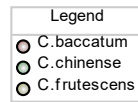
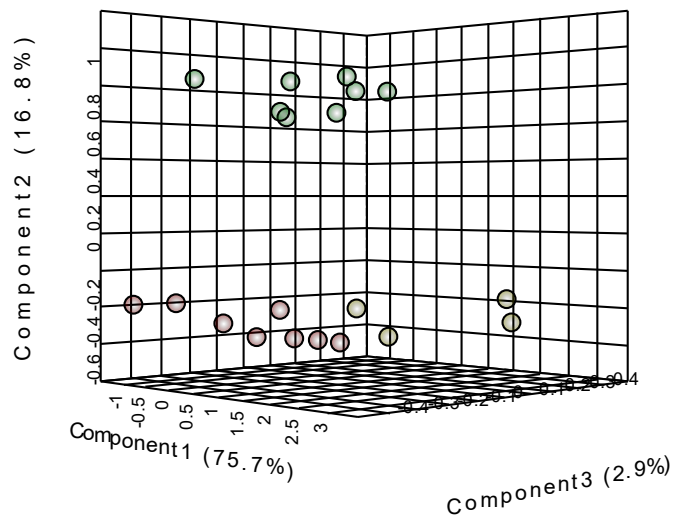


Figura 64. Diagramas de scores y loadings del modelo PLS-DA de la data sensorial por especies, $R^2 = 0.99371$, $Q^2 = 0.99094$

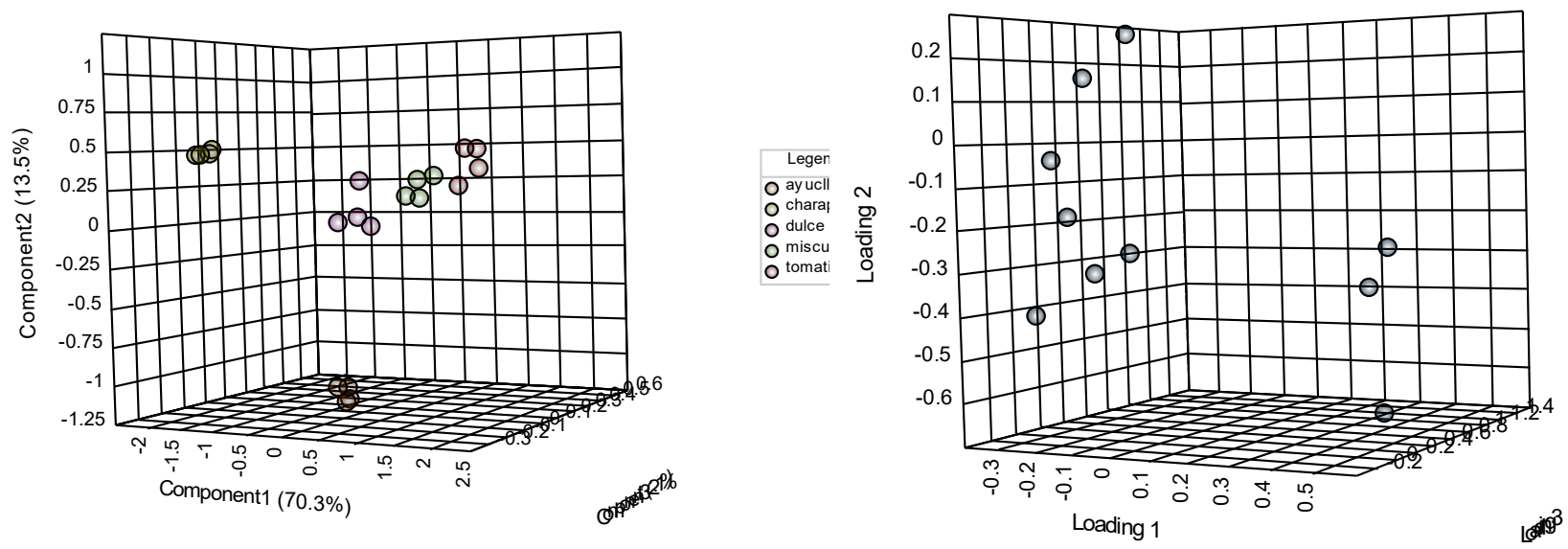


Figura 65. Diagramas de scores y loadings del modelo PLS-DA de la data sensorial por accesiones, $R^2 = 0.98409$, $Q^2 = 0.96303$

7.3.12. Perfil sensorial, diagramas de barras

El análisis de perfil sensorial mediante gráfica de barras permitió observar globalmente los descriptores que más destacan en cada accesión. Un análisis ANOVA permitió ver las diferencias significativas entre las intensidades relativas de los descriptores por accesión.

En las figuras 66-68 se observa los diagramas de barras para los 10 descriptores del sabor y aroma mostrando sus respectivas puntuaciones (score) dadas en las 5 accesiones de ajíes que conforman el perfil sensorial de los ajíes para las localidades de Chincha, Chiclayo y Huaral, respectivamente. Los diagramas fueron realizados utilizando el software GraphPad 5.0 a partir de los valores promedios de la data sensorial optimizada (4 repeticiones por descriptor para cada tipo de ají). Cada barra corresponde al valor promedio y su correspondiente barra de error (desviación estándar). Las tablas 30-32 muestran los resultados del ANOVA para los 10 descriptores por accesión de las localidades de Chincha, Chiclayo y Huaral. Un análisis Post Hoc de test de Tukey permitió hacer comparaciones múltiples de las medias. En una fila de un descriptor dado, se dan los valores promedios y su desviación estándar de la media para las 5 accesiones. En letras con diferentes superíndices (a, b, c, d) se indican los valores de las medias significativamente diferentes (existen diferencias significativas entre medias Intergrupos) ($p \leq 0.05$). También se dan los grupos con letras

de superíndice iguales que representa valores promedios no significativos (no existen diferencias significativas entre medias Intergrupos) ($p > 0.05$).

Tabla 30. ANOVA de los descriptores del sabor y aroma de Chinchá

accesión	ayuello	charapita	dulce R	miscucho	tomatito R
Orégano	3.10 ± 0.36 ^a	3.37 ± 0.14 ^a	2.85 ± 0.21 ^a	3.03 ± 0.33 ^a	2.01 ± 0.16 ^b
Hierba	5.20 ± 0.39 ^a	4.51 ± 0.43 ^a	4.66 ± 0.34 ^a	4.55 ± 0.25 ^a	3.45 ± 0.22 ^b
Manzana	4.67 ± 0.28 ^a	4.48 ± 0.11 ^{a,b}	4.01 ± 0.19 ^b	4.49 ± 0.29 ^{a,b}	3.23 ± 0.36 ^c
Cítrico	4.32 ± 0.46 ^b	5.43 ± 0.22 ^a	3.67 ± 0.29 ^{b,c}	4.27 ± 0.32 ^b	3.14 ± 0.32 ^c
Frutal	4.93 ± 0.18 ^c	6.57 ± 0.25 ^a	5.91 ± 0.10 ^b	6.07 ± 0.09 ^b	4.05 ± 0.30 ^d
Maracuyá	1.98 ± 0.15 ^{b,c}	2.93 ± 0.06 ^a	1.86 ± 0.26 ^c	2.47 ± 0.37 ^{a,b}	1.91 ± 0.39 ^{b,c}
Acidez	2.91 ± 0.15 ^a	2.65 ± 0.11 ^a	1.37 ± 0.15 ^c	1.54 ± 0.17 ^c	2.07 ± 0.19 ^b
Dulce	4.84 ± 0.14 ^b	1.92 ± 0.46 ^c	5.56 ± 0.25 ^{a,b}	4.88 ± 0.37 ^b	6.22 ± 0.48 ^a
Pimiento	5.88 ± 0.06 ^b	2.52 ± 0.25 ^c	5.97 ± 0.17 ^b	5.59 ± 0.06 ^b	6.57 ± 0.31 ^a
Tomate	3.30 ± 0.11 ^b	0.72 ± 0.25 ^e	2.53 ± 0.21 ^c	1.92 ± 0.05 ^d	3.72 ± 0.11 ^a

Tabla 31. ANOVA de los descriptores del sabor y aroma de Chiclayo

accesión	ayuello	charapita	dulce R	miscucho	tomatito R
Orégano	2.67 ± 0.24 ^{a,b}	2.94 ± 0.38 ^{a,b}	2.93 ± 0.27 ^{a,b}	2.51 ± 0.18 ^b	3.24 ± 0.38 ^a
Hierba	4.08 ± 0.17 ^b	4.85 ± 0.13 ^a	4.64 ± 0.20 ^a	3.96 ± 0.28 ^b	4.87 ± 0.17 ^a
Manzana	4.06 ± 0.28 ^b	5.09 ± 0.22 ^a	4.31 ± 0.16 ^b	4.37 ± 0.26 ^b	4.50 ± 0.52 ^{a,b}
Cítrico	4.27 ± 0.33 ^b	5.09 ± 0.46 ^a	3.77 ± 0.25 ^b	4.23 ± 0.36 ^b	3.70 ± 0.38 ^b
Frutal	4.84 ± 0.27 ^{b,c}	6.49 ± 0.41 ^a	5.23 ± 0.39 ^b	5.38 ± 0.17 ^b	4.38 ± 0.24 ^c
Maracuyá	2.24 ± 0.20 ^b	3.09 ± 0.42 ^a	1.24 ± 0.04 ^c	2.07 ± 0.19 ^b	1.86 ± 0.14 ^b
Acidez	3.14 ± 0.16 ^a	2.47 ± 0.14 ^b	1.72 ± 0.25 ^c	1.72 ± 0.11 ^c	2.31 ± 0.17 ^b
Dulce	4.13 ± 0.21 ^c	2.35 ± 0.28 ^d	5.45 ± 0.32 ^b	4.84 ± 0.38 ^b	6.19 ± 0.33 ^a
Pimiento	5.77 ± 0.12 ^b	2.75 ± 0.36 ^d	6.27 ± 0.28 ^b	4.27 ± 0.62 ^c	7.05 ± 0.13 ^a
Tomate	2.60 ± 0.22 ^b	1.28 ± 0.16 ^c	2.47 ± 0.46 ^b	1.62 ± 0.14 ^c	3.59 ± 0.22 ^a

Tabla 32. ANOVA de los descriptores del sabor y aroma de Huaral

accesión	ayuclo	charapita	dulce R	miscucho	tomatito R
Orégano	2.73± 0.24 ^b	2.97± 0.35 ^{a,b}	3.42± 0.23 ^a	2.55 ± 0.26 ^{b,c}	2.02± 0.26 ^c
Hierba	4.811± 0.32 ^a	4.24± 0.16 ^a	4.78± 0.25 ^a	4.32± 0.16 ^a	4.63± 0.37 ^a
Manzana	4.66± 0.52 ^a	4.87± 0.36 ^a	5.01± 0.23 ^a	5.03± 0.23 ^a	4.29± 0.36 ^a
Cítrico	3.89± 0.18 ^c	5.23± 0.24 ^a	4.50± 0.26 ^b	4.89± 0.16 ^{a,b}	2.92± 0.16 ^d
Frutal	4.88± 0.19 ^c	6.21± 0.24 ^a	5.96± 0.24 ^{a,b}	5.65± 0.33 ^b	4.26± 0.21 ^d
Maracuyá	1.98± 0.34 ^b	2.93± 0.17 ^a	2.43± 0.25 ^{a,b}	2.93± 0.32 ^a	1.33± 0.14 ^c
Acidez	2.59± 0.39 ^a	1.77± 0.22 ^c	1.65± 0.22 ^c	1.94± 0.22 ^{b,c}	2.38± 0.14 ^{a,b}
Dulce	3.39± 0.57 ^b	2.08± 0.55 ^c	4.30± 0.28 ^a	2.83± 0.21 ^{b,c}	4.78± 0.32 ^a
Pimiento	4.56± 0.40 ^a	1.94± 0.27 ^c	4.63± 0.24 ^a	3.43± 0.29 ^b	5.19 ± 0.55 ^a
Tomate	1.74 ± 0.30 ^b	0.82± 0.15 ^c	1.51± 0.05 ^b	1.40± 0.20 ^{b,c}	2.40± 0.47 ^a

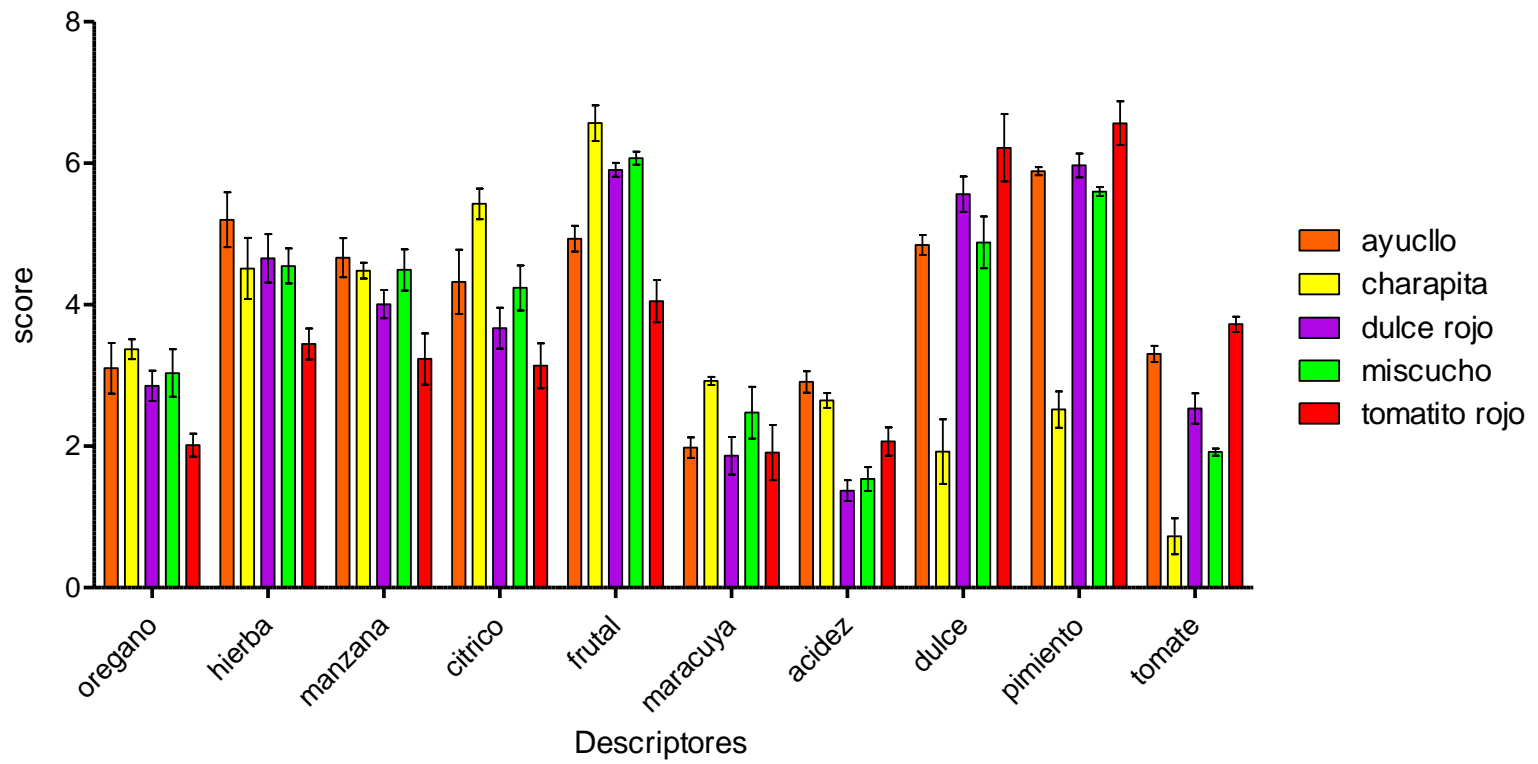


Figura 66. Diagrama de barras sensorial de los descriptores de la localidad de Chincha

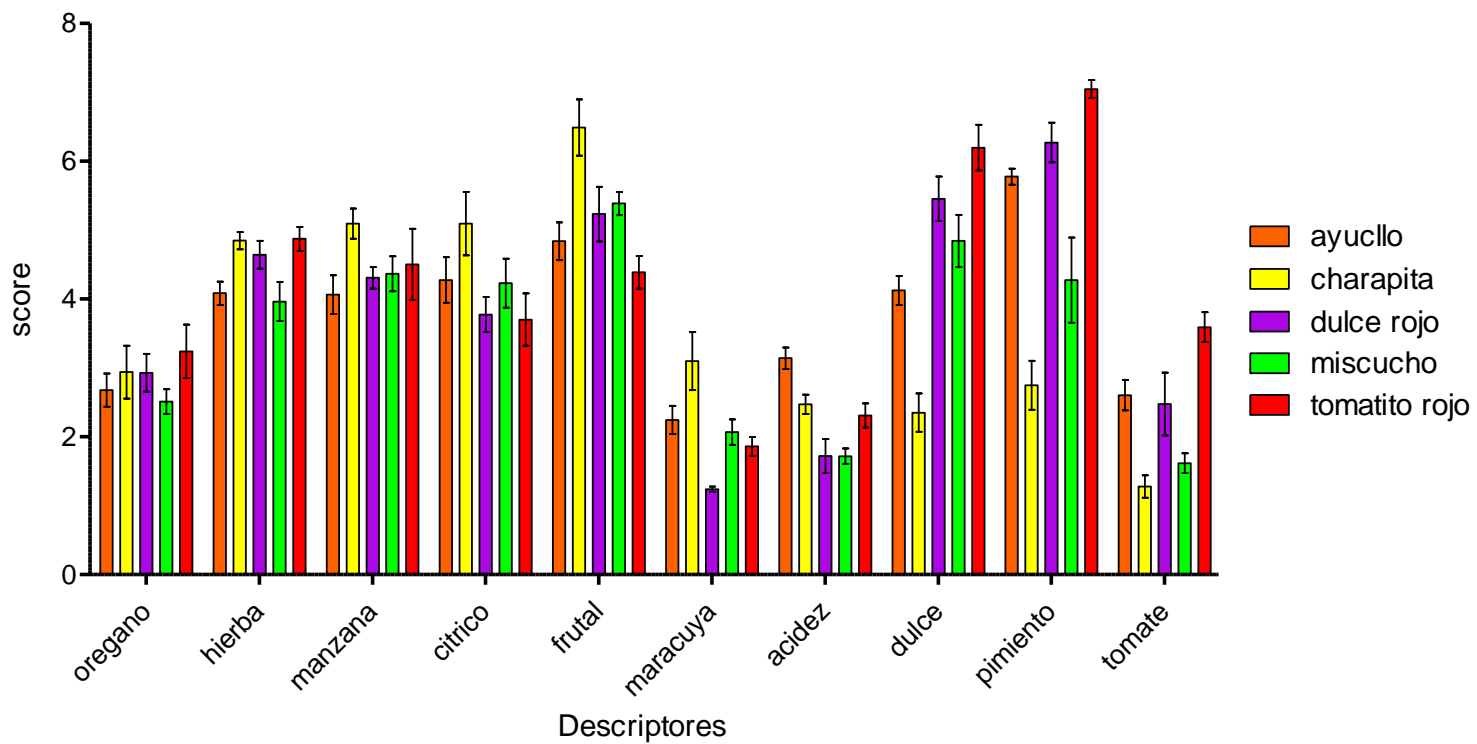


Figura 67. Diagrama de barras sensorial de los descriptores de la localidad de Chiclayo

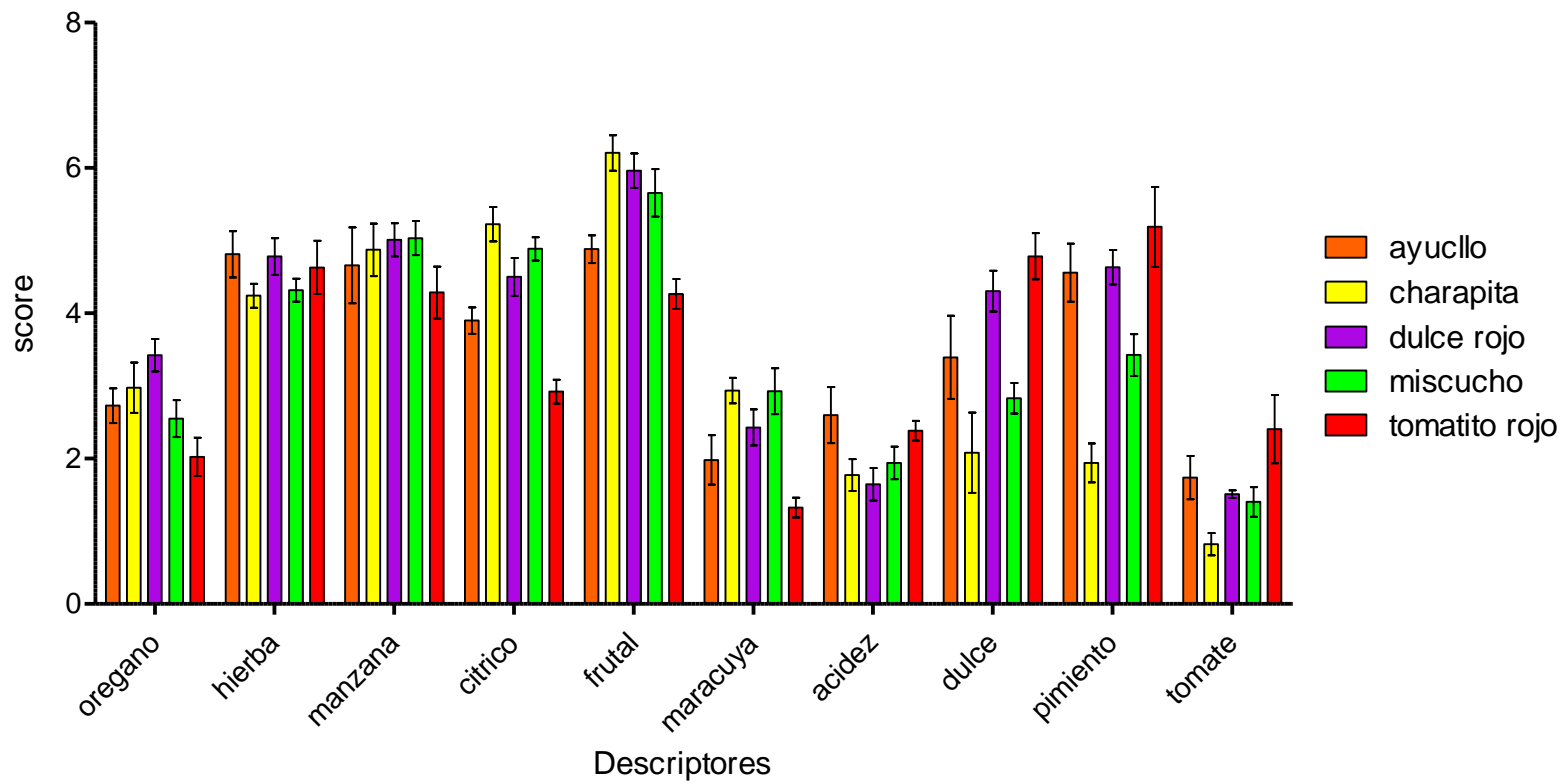


Figura 68. Diagrama de barras sensorial de los descriptores de la localidad de Huaral

7.4. Análisis multivariado de metabolitos volátiles

El objetivo del presente análisis fue encontrar el mejor modelo estadístico que explique las variables de metabolitos volátiles con sus correspondencias categóricas tanto como especies y accesiones. Los metabolitos volátiles fueron analizados en las muestras de ajíes frescos y liofilizados. Para los ajíes frescos se utilizó una fibra SPME 85 μm celeste compuesta de un revestimiento absorbente de poliacilato Carboxen/Polimetilsiloxano (CAR/PDMS), con capacidad de extraer metabolitos semivolátiles en el rango de masa molecular MW 80-300 uma. Mientras que con los ajíes liofilizados se utilizó una fibra SPME 50/30 μm ploma compuesta de un revestimiento absorbente de Divinilbenceno/Carboxen/Polimetilsiloxano (DVB/CAR/PDMS), exclusivo para aromas (metabolitos volátiles y semivolátiles) en el rango de masa molecular MW 40-275 uma. Las figuras 69 y 70 muestran los resultados del análisis multivariado de PCA y OPLS-DA por especies de la data cromatográfica de Chincha de las fibras SPME celeste y ploma en escalado pareto, respectivamente. Las figuras 71 y 72 muestran los resultados del PCA y OPLS-DA de las dos tablas de datos cromatográficos alineados y combinados (fibras SPME celeste y ploma) en escalado pareto para los análisis por especies y accesiones. Como se pudo observar en los cromatogramas GC-MS (Figuras 35-39) de las fibras SPME celeste y ploma, ambos perfiles cromatográficos son complementarios. Para mejorar el modelo, se procedió con la normalización de la data GC-MS combinada. Para ello se ensayaron 20 métodos de normalización utilizando el software Online NOREVA (57), de los cuales, el método EigenMS fue el que tuvo mejor

performance. El método de normalización Eigen permitió eliminar la variación intraclase e interclase causadas por variables aleatorias y experimentales.

En la figura 73 se observa los resultados de los diagramas boxplots de la data GC-MS combinada antes y después de la normalización por el método EigenMS. En las figuras 74 y 75 se observan los resultados del PCA del software NOREVA que se utilizaron como primer criterio para evaluar la performance de la normalización EigenMS. Se observó una mejora en el criterio de clasificación, así como un aumento en la varianza explicada en el modelo final. El segundo criterio que utilizó el software NOREVA para evaluar la performance del método de normalización EigenMS fue el área bajo la curva (AUC) de los gráficos ROC (Receiver Operating Characteristics) que se muestra en la figura 76, que se utilizó para evaluar el rendimiento de la clasificación (accesiones o especies) donde el valor del área indica la precisión de la clasificación. El AUC es una de las medidas de evaluación más importantes para verificar el rendimiento de cualquier modelo de clasificación. El área bajo la curva AUC de la gráfica ROC representa el grado o medida de separabilidad e indica cuánto del modelo es capaz de distinguir entre las clases. Cuanto mayor sea el AUC, mejor será el modelo para predecir las categóricas (accesiones o especies). Un modelo excelente tiene AUC cerca del 1, lo que significa que el modelo ofrece una clasificación interclase más precisa (mejor separación interclase). Un modelo pobre tiene AUC cerca de 0, lo que significa que tiene la peor medida de separabilidad. Es decir, predice 0s como 1s y 1s como 0s (ceros y unos representan a las diferentes variables categóricas). Y cuando AUC es

0.5, significa que el modelo no tiene capacidad de separación de clases. Los resultados de la figura 72 muestran que el mejor modelo lo da la normalización EigenMS por accesiones AUC= 1. Las figuras 77 y 78 muestran los gráficos de PCA y OPLS-DA de la data GC-MS normalizada por el método EigenMS por especies y accesiones utilizando el software SIMCA-P 15. En los gráficos de PCA los valores de R^2 y Q^2 nos indican la calidad del modelo. Las variables categóricas se ajustan mejor al modelo del análisis por accesiones ($R^2 = 0.916$) y tiene mejor capacidad predictiva ($Q^2 = 0.838$) que el modelo del análisis por especies (menores valores de R^2 y Q^2).

Las figuras 79 y 80 muestran los diagramas Heatmap de la data GC-MS normalizada por EigenMS de accesiones y especies, construidas con el software online MetaboAnalyst 3.0. Se observan claras diferencias entre especies y accesiones debido a diferencias en intensidades entre bloques de metabolitos respectivamente.

En la tabla 33 se dan el número de metabolitos de cada clase química identificado en los ajíes frescos y liofilizados (tablas combinadas). Se han identificado 12 principales clases químicas entre las cuales destacan ésteres, monoterpenos, sesquiterpenos y compuestos alifáticos como los metabolitos volátiles con mayor abundancia en los ajíes nativos.

Tabla 33. Principales Clases químicas de los metabolitos volátiles, ajíes frescos y liofilizados

#	Clase química	% Clases químicas
1	Acido	2.8
2	Alcohol	8.3
3	Aldehído	4.1
4	Alifáticos	11.1
5	Aromático	1.8
6	Cetonas	4.9
7	Ester	35.0
8	Éter	2.8
9	Furano	0.8
10	Lactona	2.6
11	Monoterpenos	10.1
12	Sesquiterpenos	9.8
13	Otros	5.7

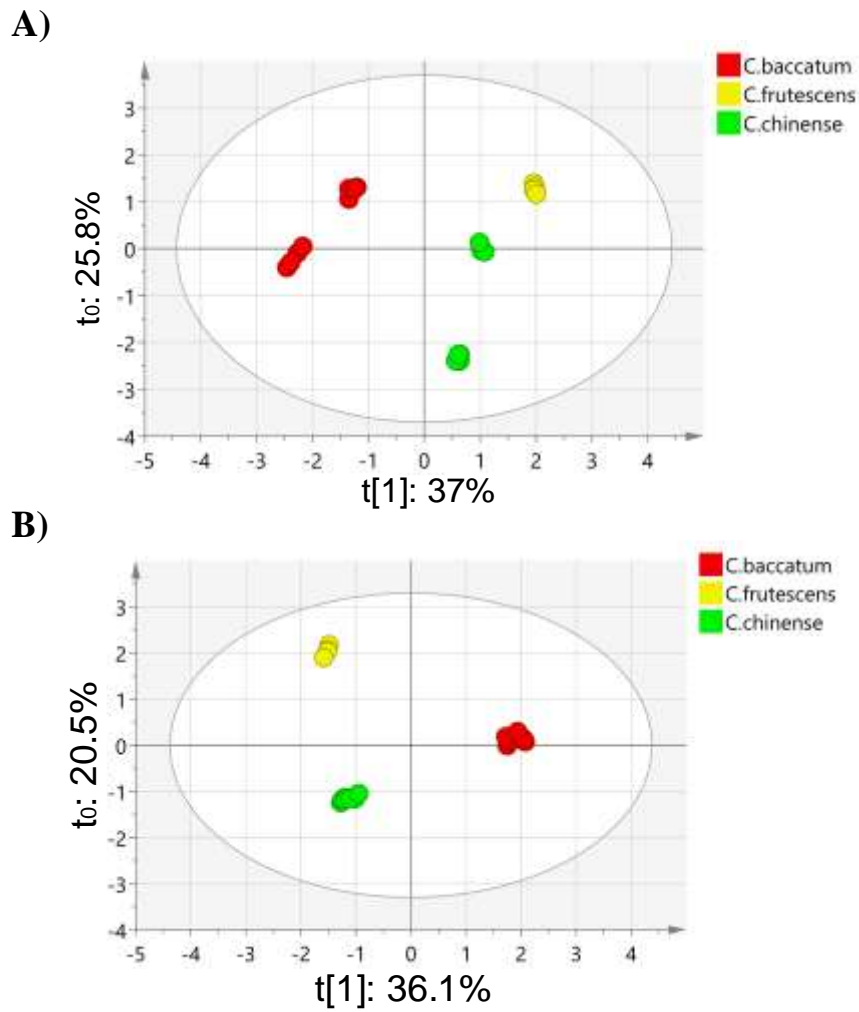


Figura 69. PCA(A) y OPLS-DA (B) de los perfiles GC-MS de las tres localidades de Chíncha, Chiclayo y Huaral, clase especie, fibra SPME celeste, escalado pareto. El PCA muestra un $R^2X=0.928$, el OPLS-DA muestra un $R^2Y=0.995$ y $Q^2=0.998$. CV-ANOVA ($p=3.17921e-23$). La varianza explicada es del 62.8 y 56.6%, respectivamente.

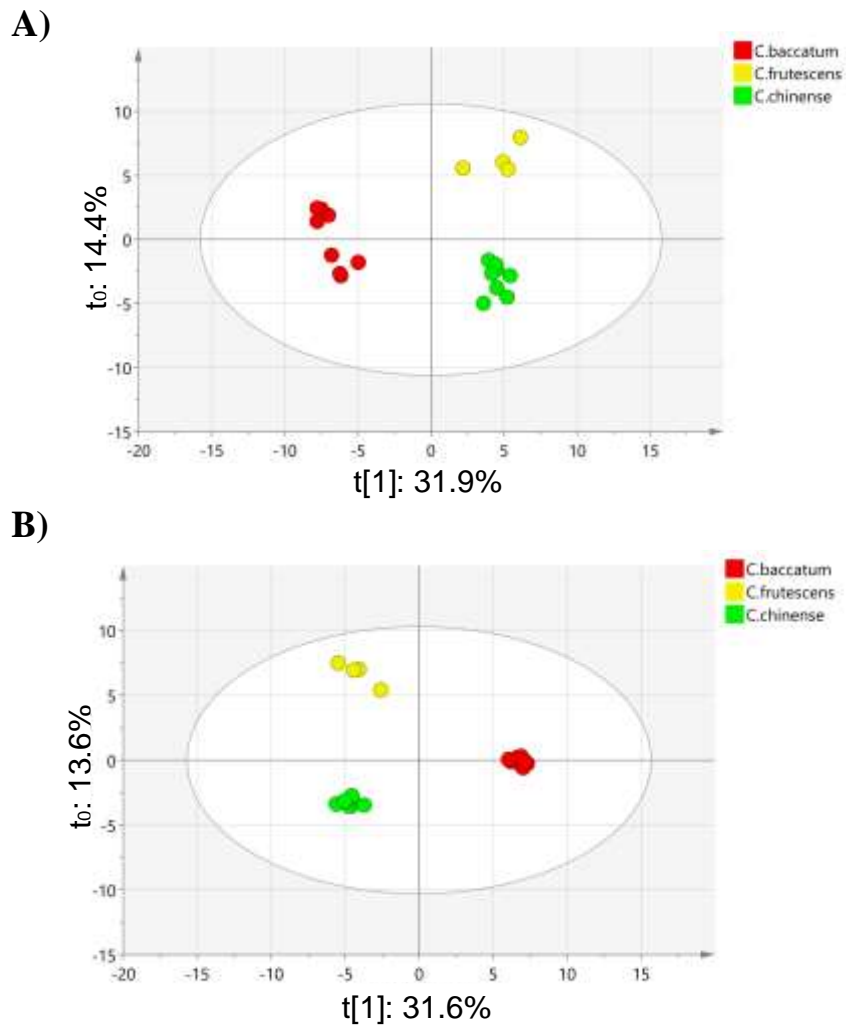


Figura 70. PCA(A) y OPLS-DA (B) de los perfiles GC-MS de la localidad de Chíncha, clase especie, fibra SPME ploma, transformación logarítmica, escalado pareto. El PCA muestra un $R^2X= 0.576$, el OPLS-DA muestra un $R^2Y= 0.987$ y $Q^2 = 0.936$. CV-ANOVA ($p = 4.14157e-12$). La varianza explicada es del 46.3 y 45.2%, respectivamente.

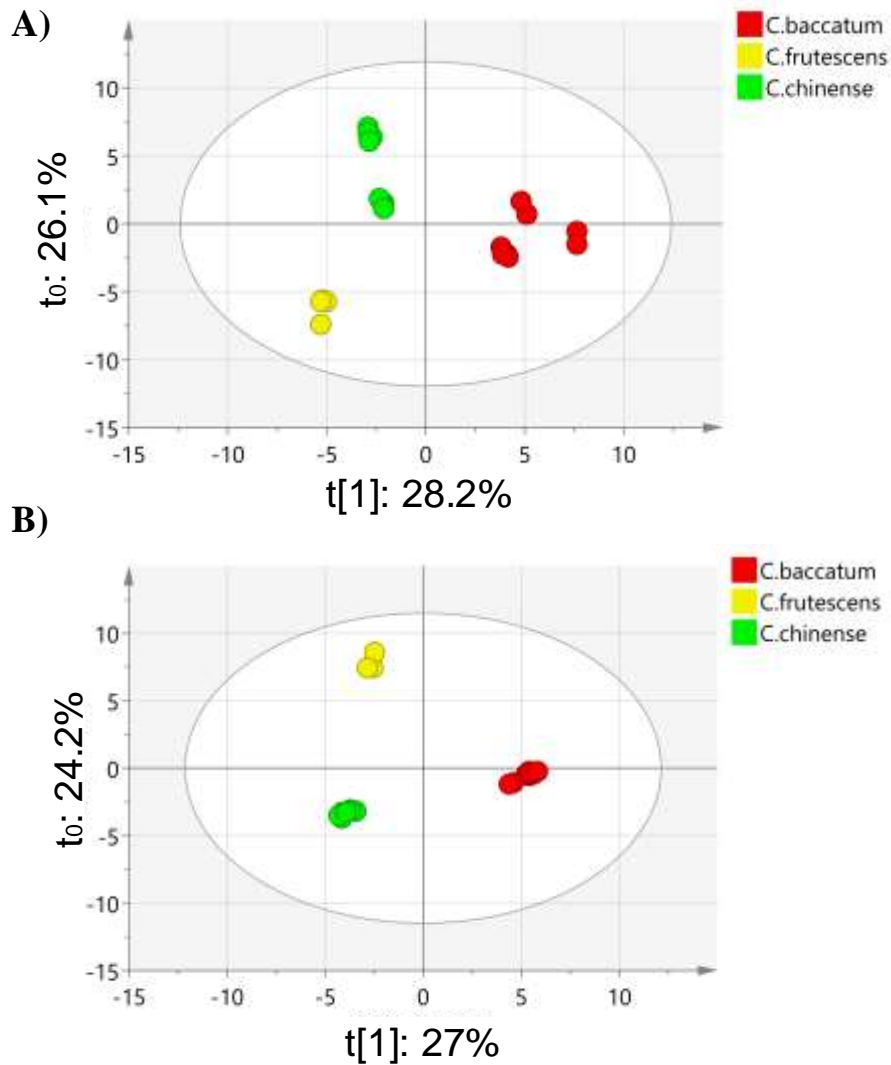


Figura 71. PCA(A) y OPLS-DA (B) de los perfiles GC-MS de la localidad de Chíncha, tablas combinadas de datos normalizados (fibra celeste y pluma), clase especie, escalado pareto. El PCA muestra un $R^2X= 0.814$, el OPLS-DA muestra un $R^2Y= 0.993$ y $Q^2 = 0.978$. CV-ANOVA ($p = 7.49841e-18$). La varianza explicada es del 54.3 y 51%, respectivamente.

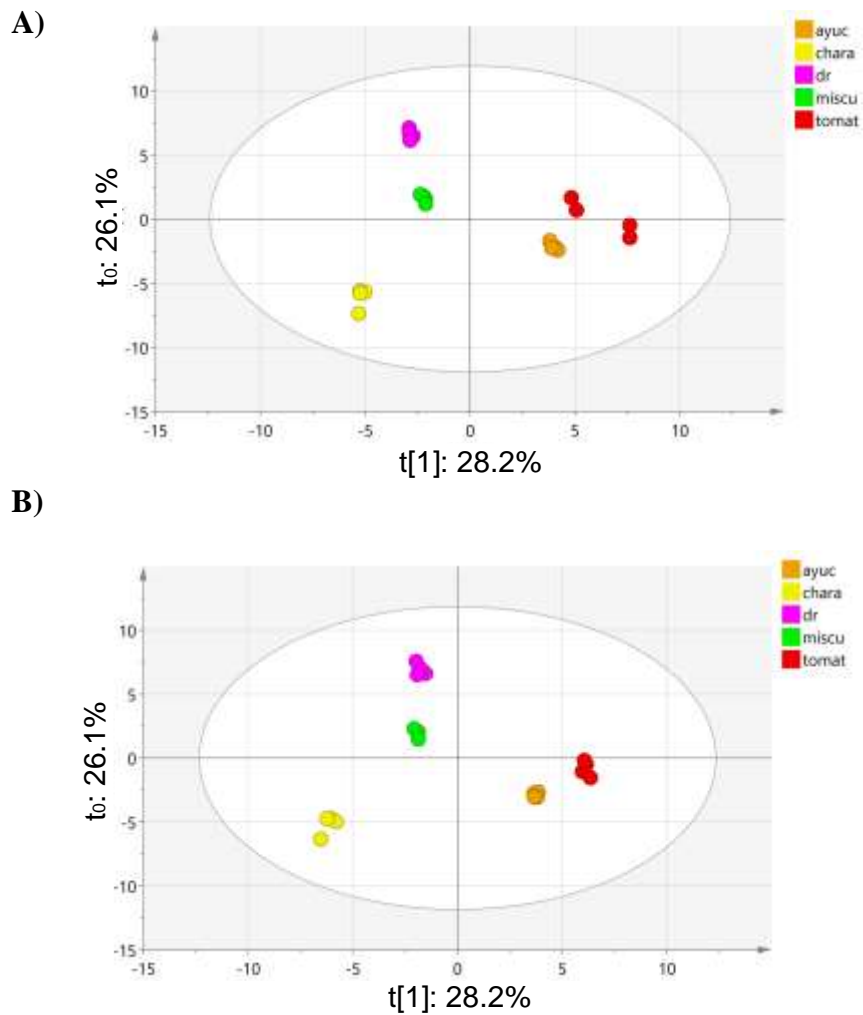


Figura 72. PCA(A) y OPLS-DA (B) de los perfiles GC-MS de la localidad de Chincha, tablas combinadas de datos normalizados (fibra celeste y ploma), clase accesión. El PCA muestra un $R^2X=0.814$, el OPLS-DA muestra un $R^2Y=0.99$ y $Q^2=0.956$. CV-ANOVA ($p=1.60108e-22$). La varianza explicada es del 54.3%, respectivamente.

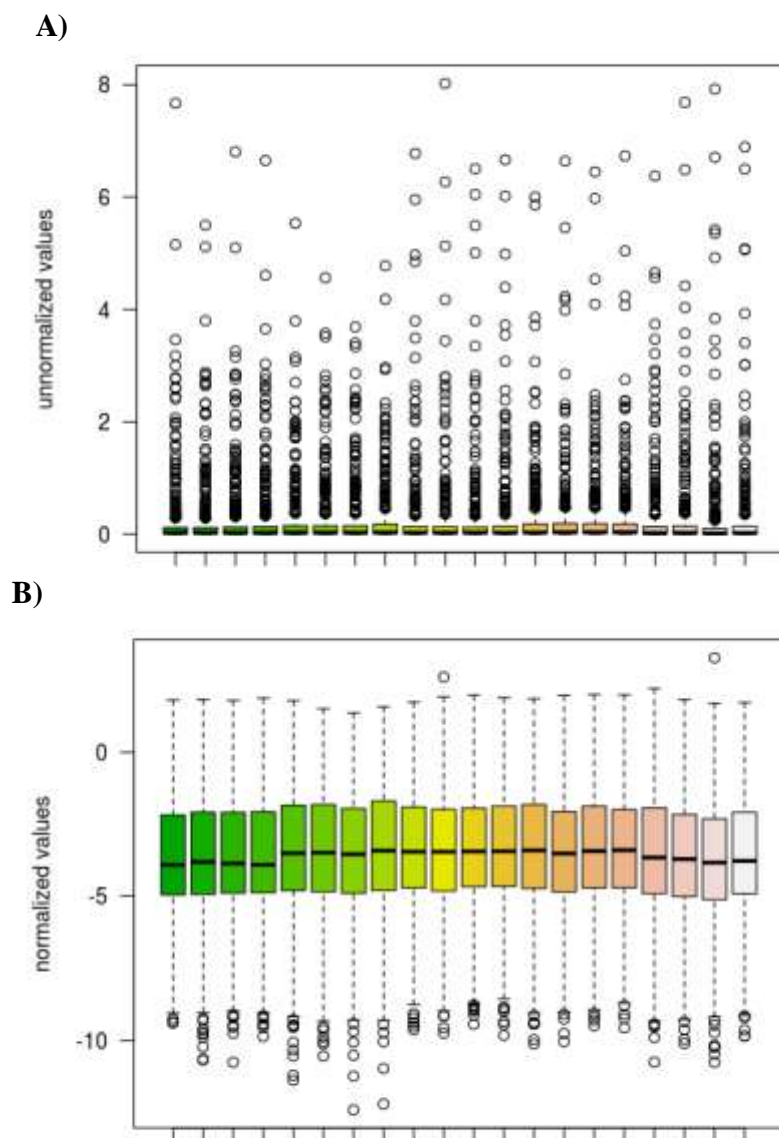


Figura 73. Diagramas box plots de la data GC-MS combinada antes (A) y después (B) de la normalización por el método EigenMS, clase especie.

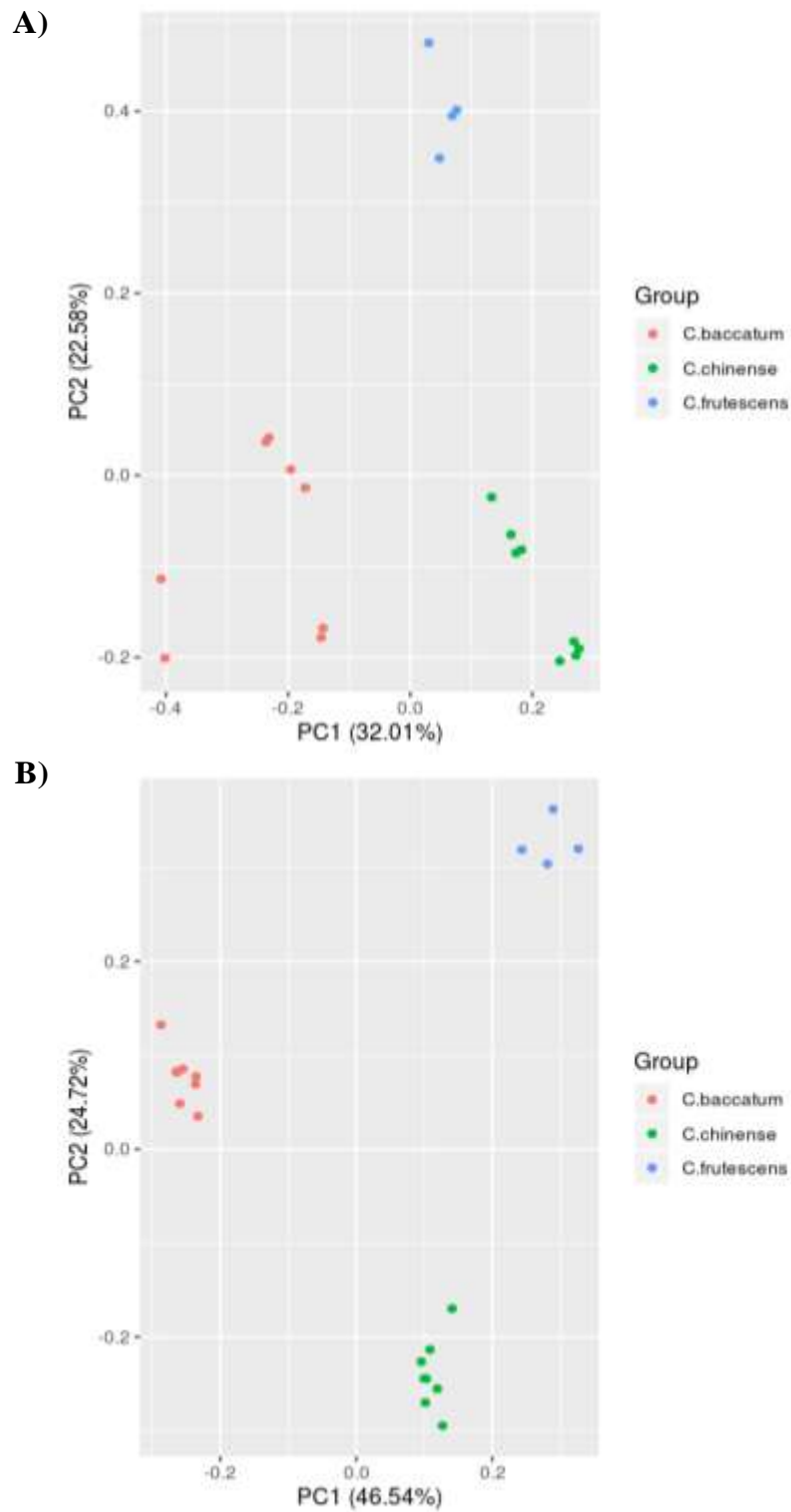
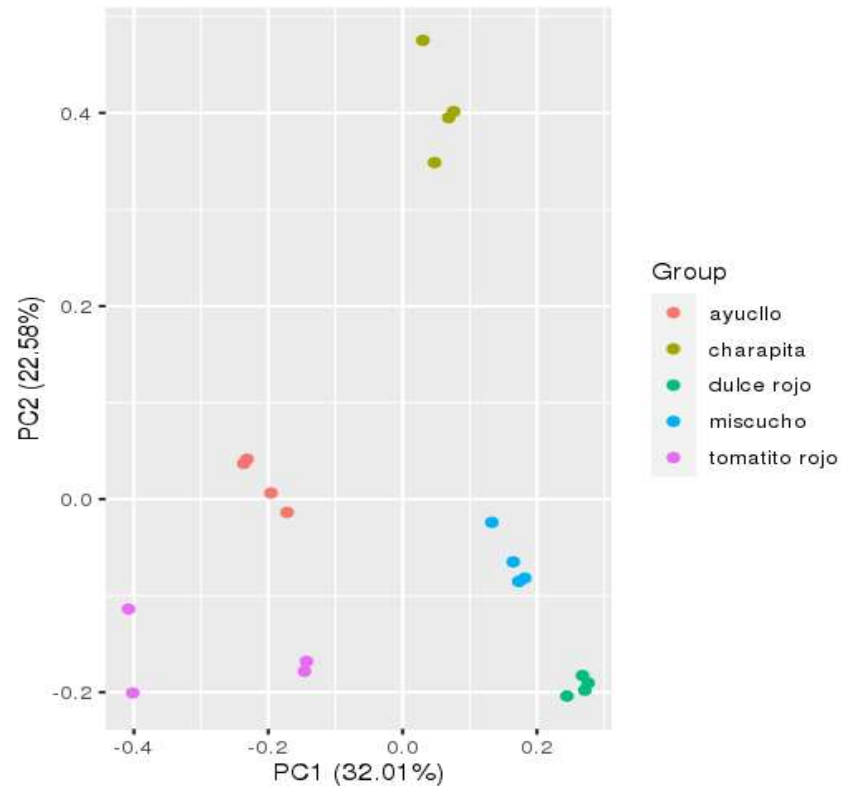


Figura 74. Evaluación del performance de la normalización por especie, se muestra el PCA antes (A) y después (B) de la normalización por el método EigenMS (varianza explicada 75.84%).

A)



B)

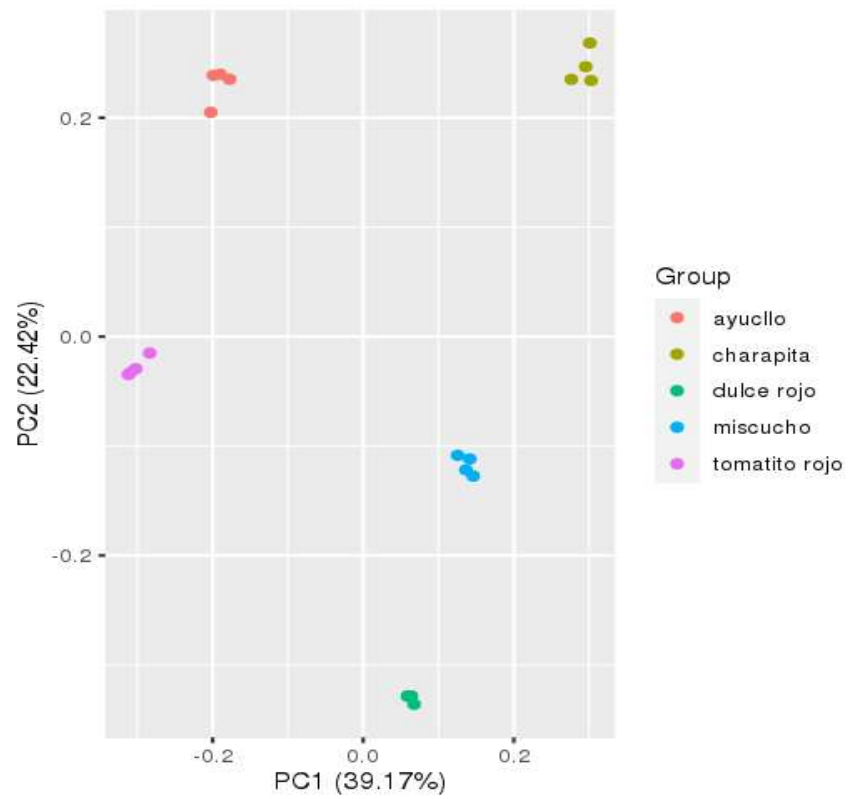
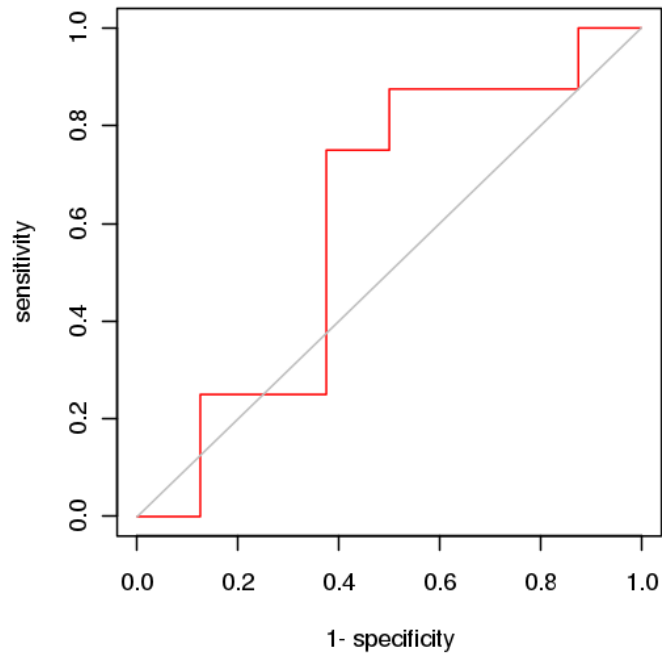


Figura 75. Evaluación del performance de la normalización por accesión, se muestra el PCA antes (A) y después (B) de la normalización por el método EigenMS (varianza explicada 61.59%).

A)



B)

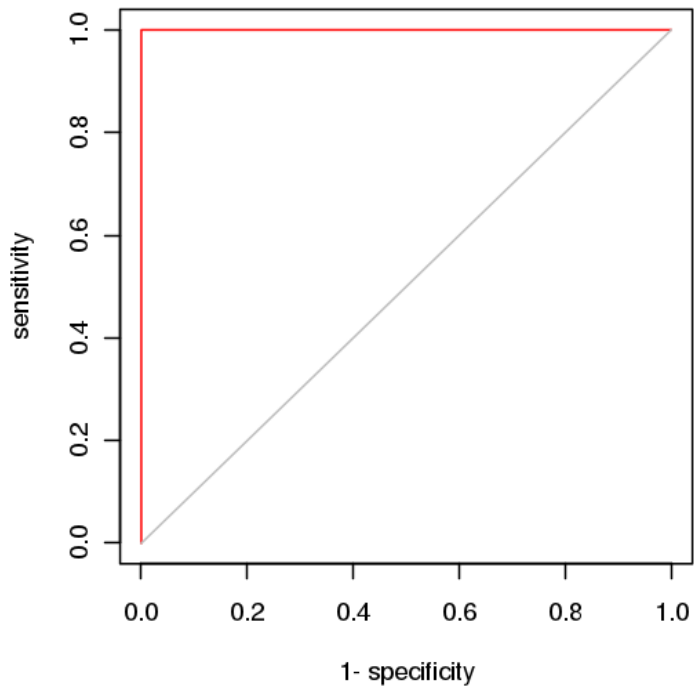


Figura 76. Gráficos ROC para la evaluación de la influencia del método de normalización EigenMS en la precisión de la clasificación. Para la clasificación por especies se observa un Área bajo la curva (AUC): AUC = 0.6406 por especies, y por accesiones un AUC = 1. Un AOC > 0.7 (CV diez veces). Se considera como buena precisión en la clasificación.

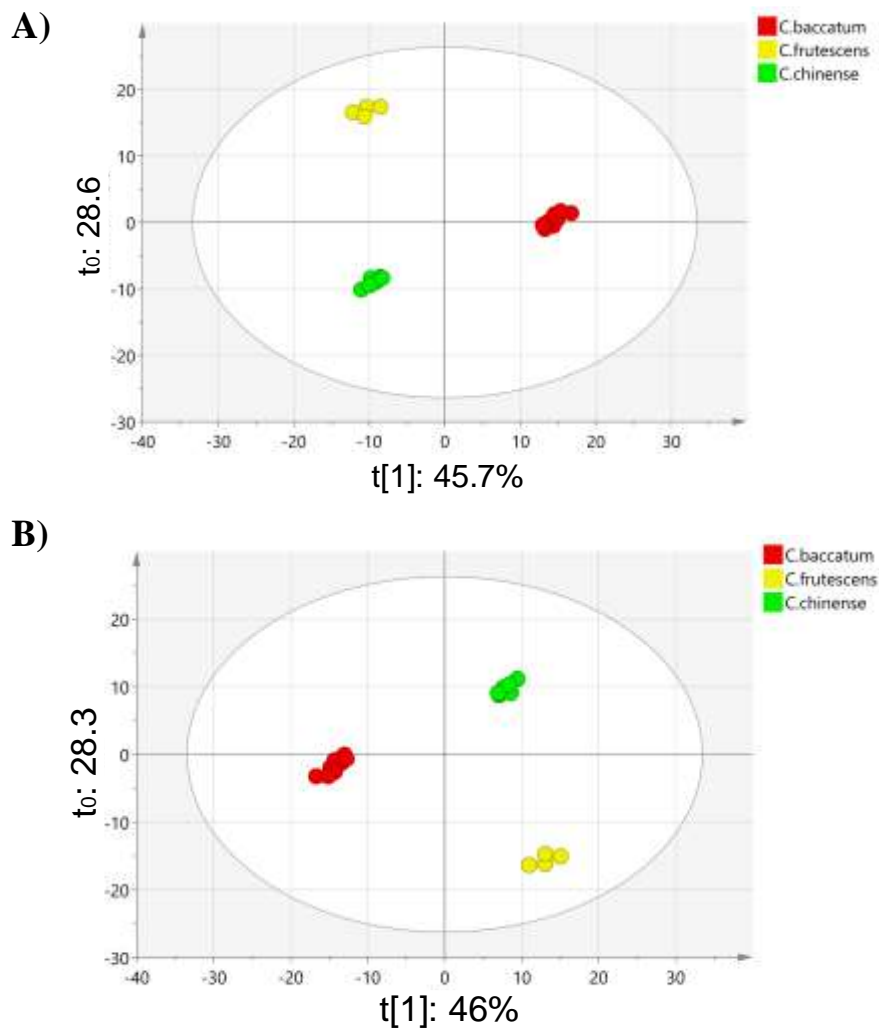


Figura 77. PCA(A) y OPLS-DA (B) de los perfiles GC-MS, localidad de Chincha, normalizada por el método EigenMS, clase especie. El PCA muestra un $R^2X= 0.743$, el OPLS-DA muestra un $R^2Y= 0.992$ y $Q^2 = 0.985$. CV-ANOVA ($p = 6.30892e-27$). La varianza explicada es del 74.3% respectivamente.

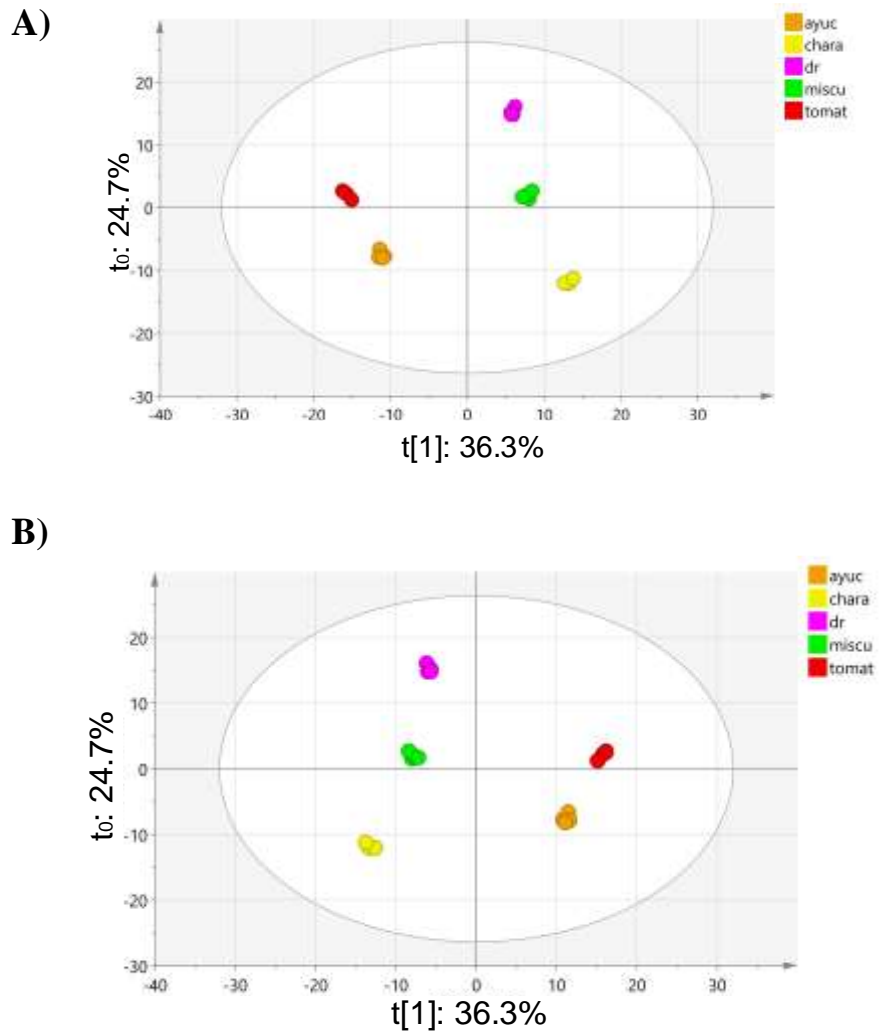


Figura 78. PCA(A) y OPLS-DA (B) de los perfiles GC-MS, localidad de Chincha, normalizada por el método EigenMS, clase accesión. El PCA muestra un $R^2X=0.916$, el OPLS-DA muestra un $R^2Y=0.997$ y $Q^2=0.993$. CV-ANOVA ($p=0$). La varianza explicada es del 61%, respectivamente.

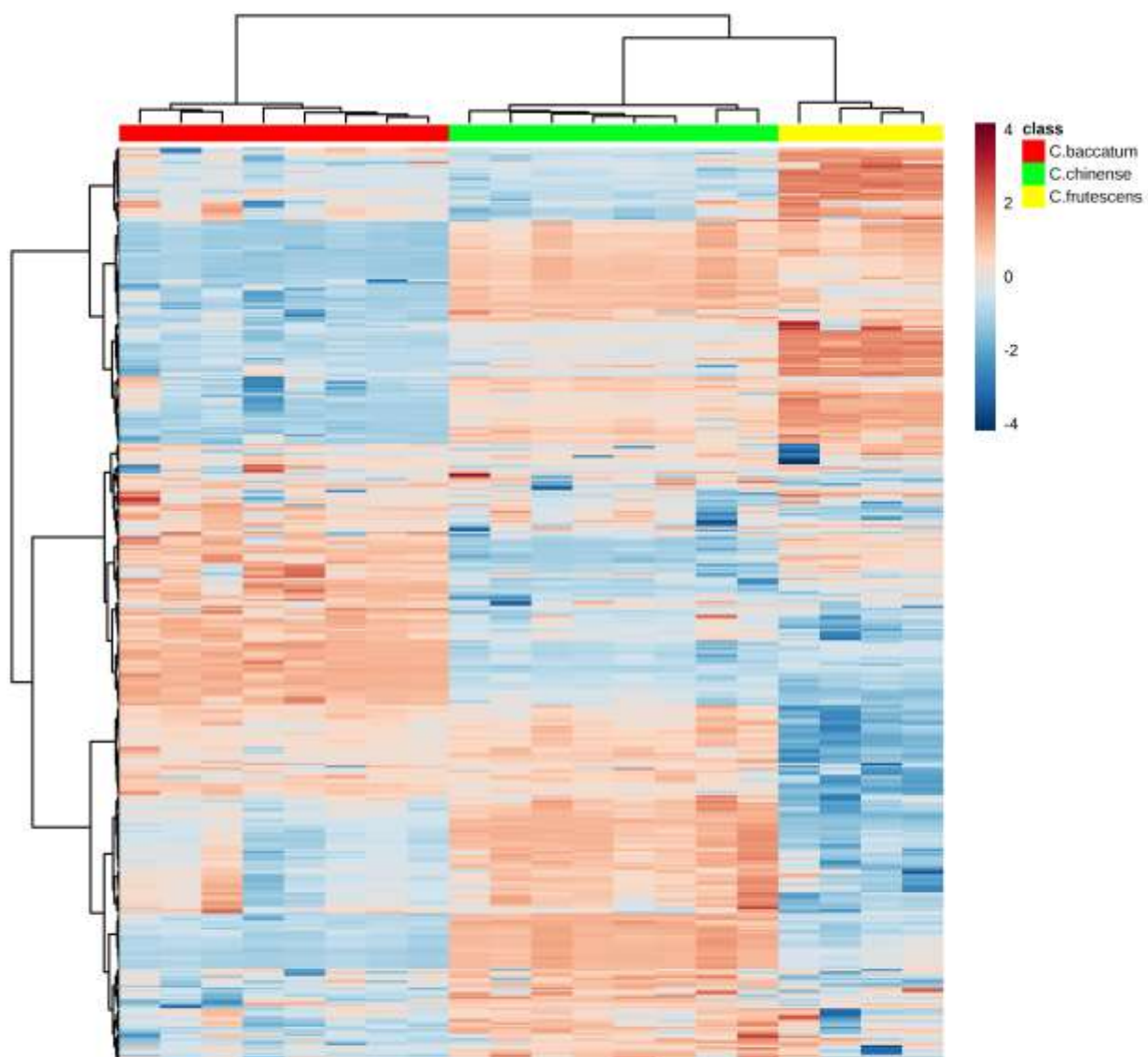


Figura 79. Heatmap de las 397 características de metabolitos volátiles identificados a partir del modelo PLS-DA de la clase especie de ajés de la localidad de Chincha.

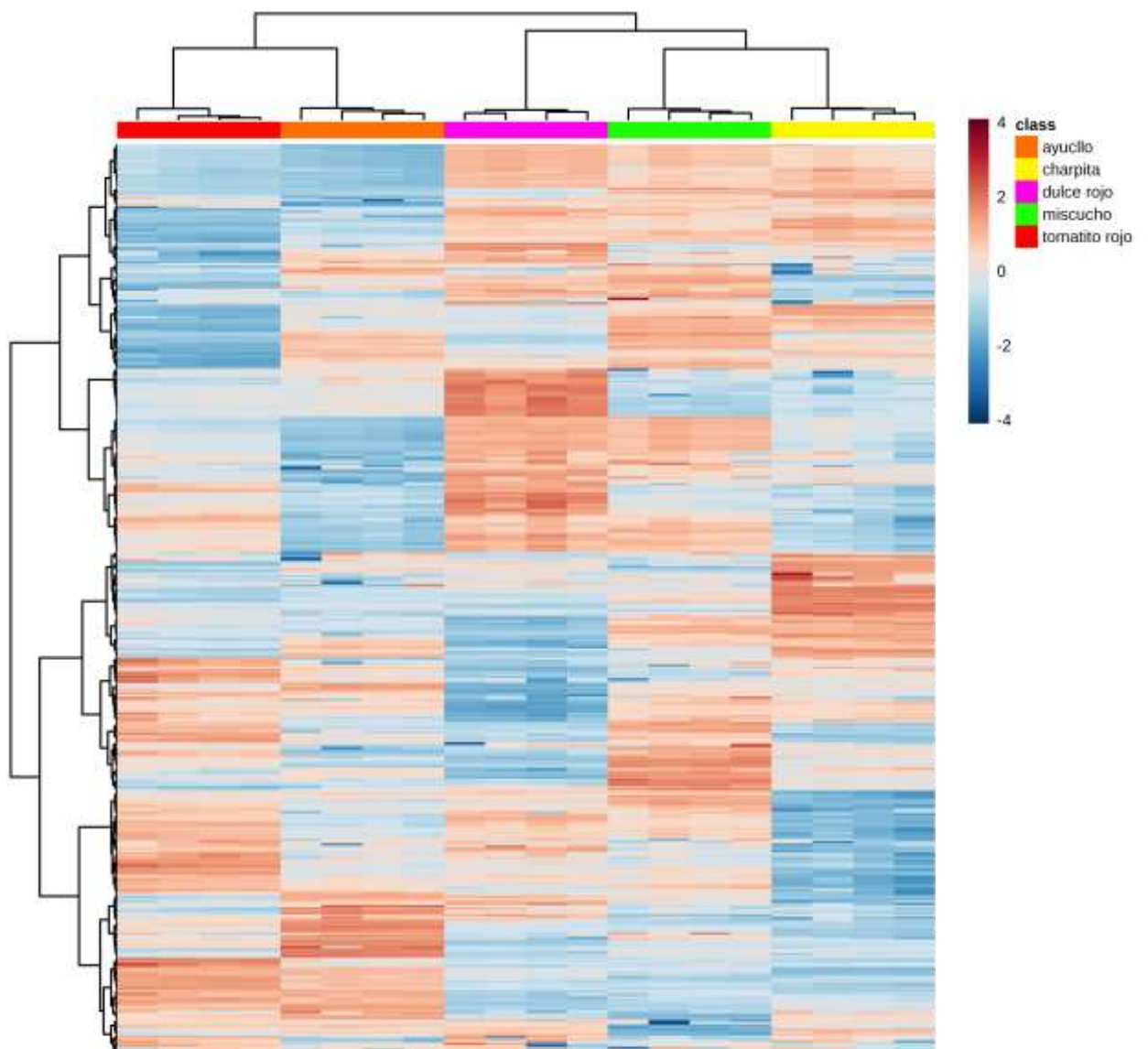


Figura 80. Heatmap de las 397 características de metabolitos volátiles identificados a partir del modelo PLS-DA de la clase accesión de ajíes de la localidad de Chíncha.

7.5. Diferencias Metabolómicas entre cosecha Orgánica y Convencional

La figura 81, 82 y 83 muestra el PCA (A) de la data cromatográfica de los ajíes de las especies (*C. frutescens*), *C. baccatum* y *C. chinense* de las cosechas orgánica y convencional de las tres localidades: Chincha (cosecha orgánica), Chiclayo y Huaral (cosecha Convencional) respectivamente. El Análisis PCA mostró una diferenciación entre ambas cosechas (se muestra en líneas punteadas en azul), mientras que el OPLS-DA (B) mostró una mejor discriminación y separación de clases (Orgánica vs Convencional). El análisis S-plot (C) puso en evidencia los marcadores más destacados para ambas cosechas. Las tablas 34, 35 y 36 muestran los resultados del análisis S-plot para las tres especies. El análisis Boxplots (figura 84) de la lista de metabolitos de las tablas 34, 35 y 36 mostró que sólo hubo dos metabolitos marcadores (4-hydroxi-2-pyrrolidincarboxylic acid; N-Me y Capsaicina) para la discriminación entre ambas cosechas orgánica y convencional tanto por especies y accesiones. El análisis Boxplots para la accesión charapita mostró que el (S)-malato es discriminante para la cosecha orgánica y el ácido cítrico para la cosecha convencional. Para la accesión ayuclo se encontraron dos marcadores discriminantes de la cosecha convencional, el ácido cítrico y el compuesto 6-O- α -D-Glucopyranuronosyl-D-glucose; α -pyranose-form. De la misma forma estos dos metabolitos mencionados para la accesión ayuclo fueron marcadores para la accesión tomatito rojo de la cosecha orgánica. Para la accesión dulce rojo se encontraron dos metabolitos marcadores, el metabolito 2,3-Dihydro-5,7-dihydroxy-2-methyl-4H-1-benzopyran-4-one fue discriminante para la cosecha

orgánica y Azadirachnol para la cosecha convencional. Estos dos últimos metabolitos mencionados no fueron específicos para la cosecha orgánica y convencional en la accesión miscucho.

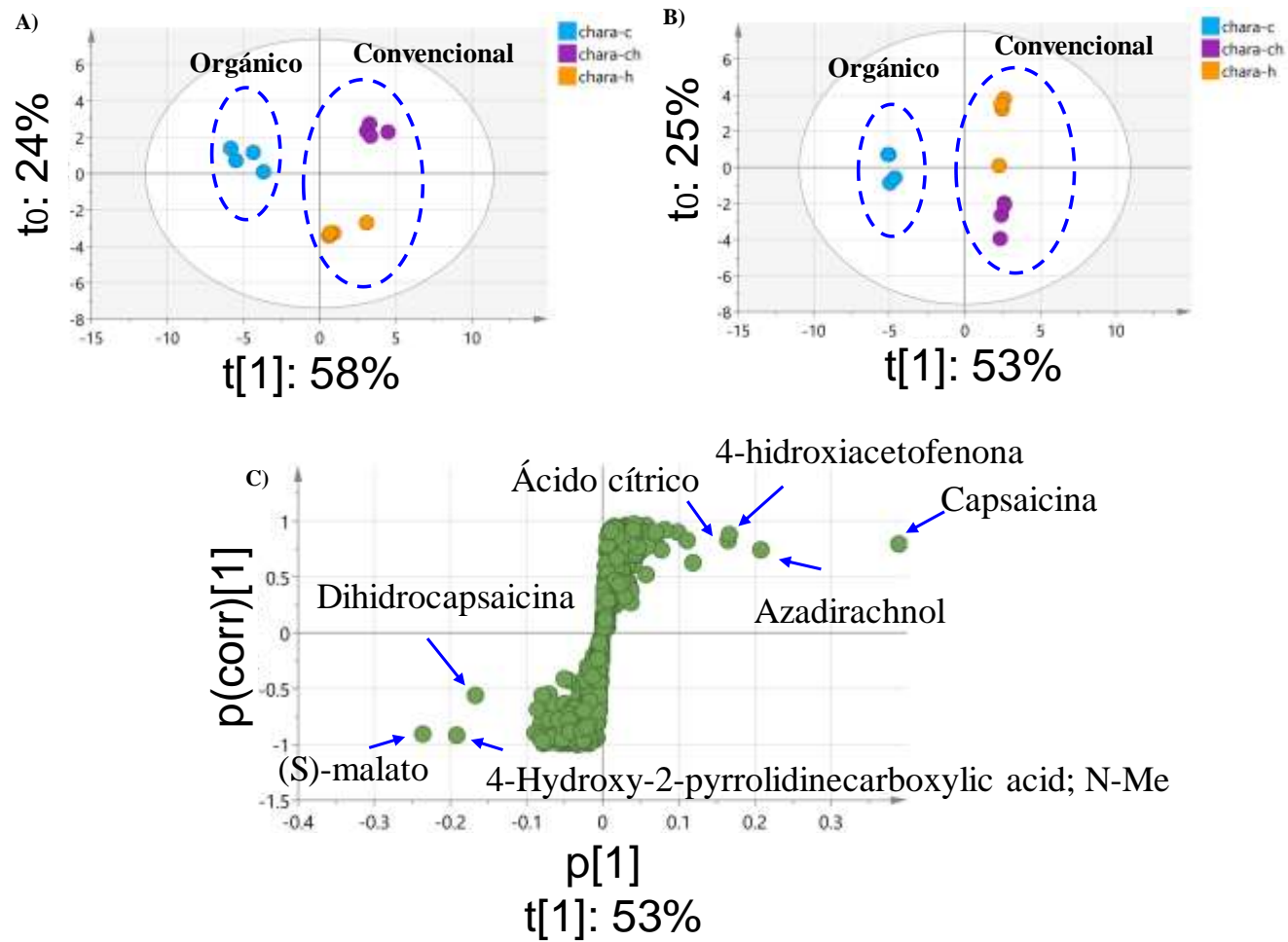


Figura 81. Perfil Metabólico de la cosecha de *C. frutescens* Orgánico vs Convencional. PCA(A) $R^2=0.93$, OPLS-DA (B) $R^2 = 0.998$, $Q^2 = 0.985$ y S-plot (C), se indican los principales Marcadores entre ambas cosechas. Leyenda, chara = charapita, c= chincha, ch= chiclayo, c= hualal.

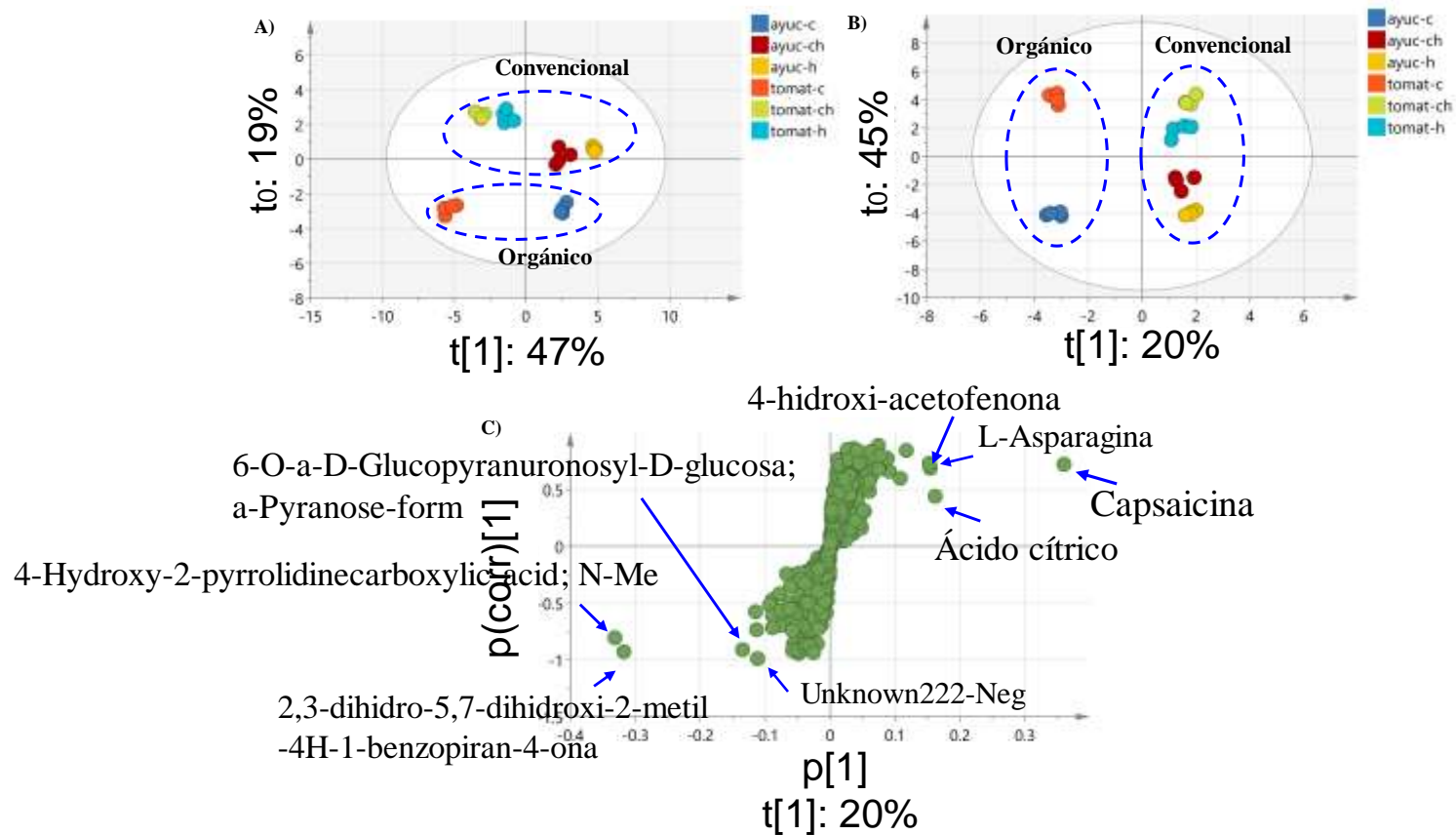


Figura 82. Perfil Metabolómico de la cosecha de *C. bacatum* Orgánico vs Convencional. PCA(A) $R^2=0.883$, OPLS-DA (B) $R^2=0.987$, $Q^2=0.971$ y S-plot (C), se indican los principales Marcadores entre ambas cosechas. Leyenda, ayuc= ayucello, tomat= tomatito rojo.

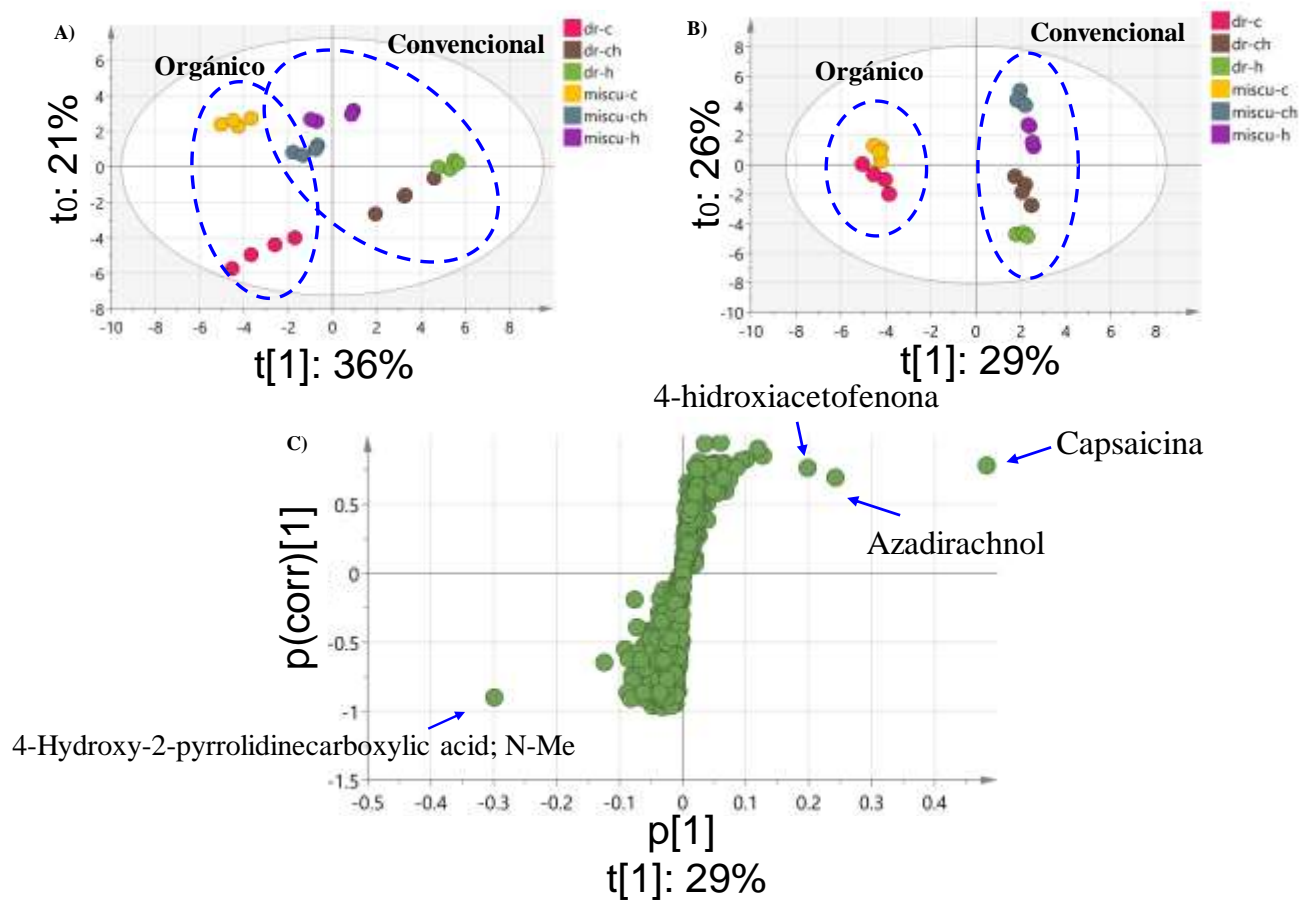


Figura 83. Perfil Metabolómico de la cosecha de *C. chinense* Orgánico vs Convencional. PCA(A) $R^2=0.995$, OPLS-DA (B) $R^2 =0.992$, $Q^2 = 0.978$ y S-plot (C), se indican los principales Marcadores entre ambas cosechas. Leyenda: dr = dulce rojo, miscu = miscucho.

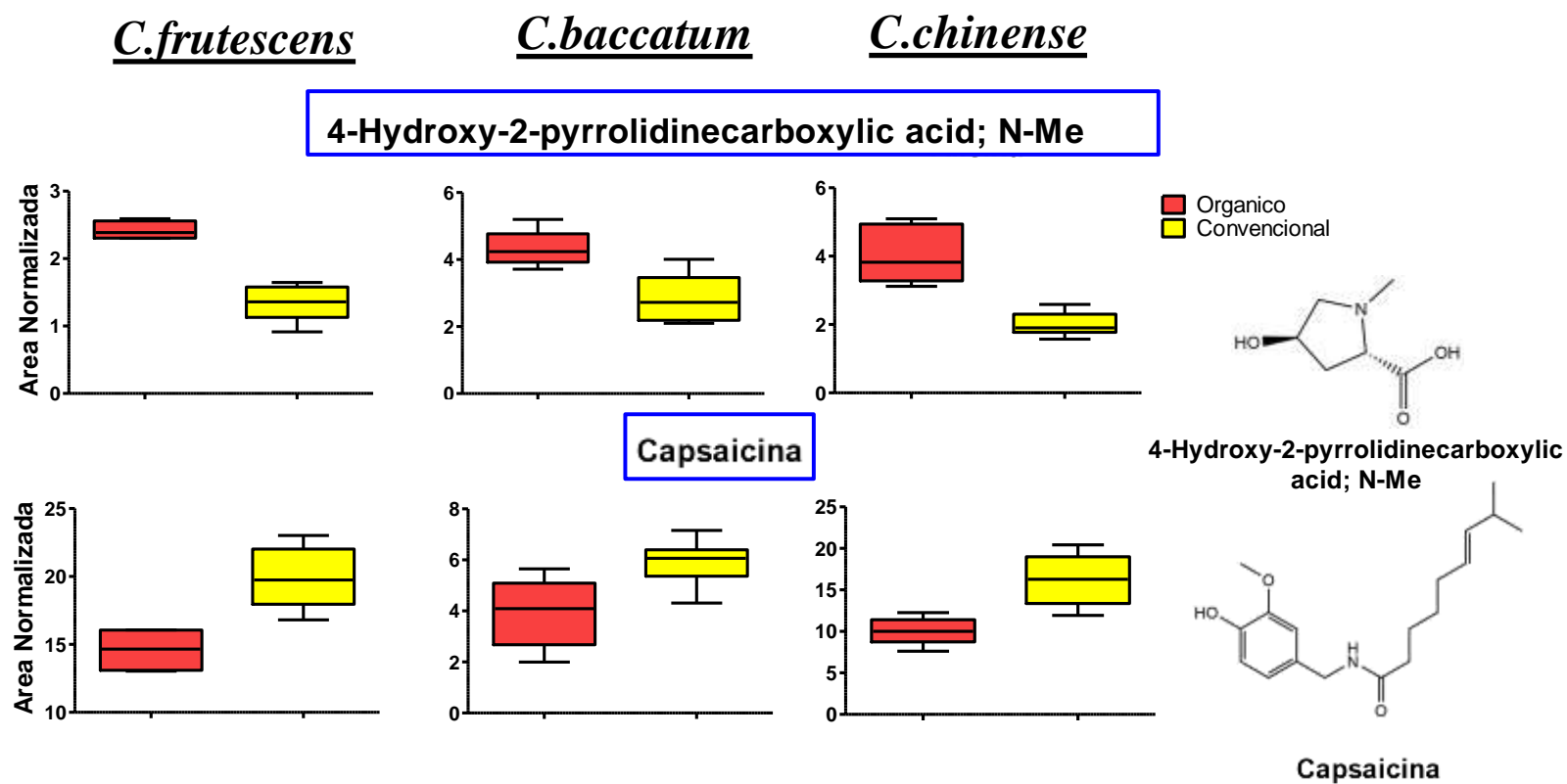


Figura 84. Diagramas Boxplots de áreas cromatográficas relativa normalizada de 4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me y Capsaicina, marcadores discriminantes de cosechas Orgánica y Convencional.

Tabla 34. Metabolitos Marcadores de *C. frutescens* Orgánico vs Convencional por análisis S-plot

Cosecha	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
Orgánico	Neg	0.74	133.01	C4H4O5	(S)-Malate	Carbohidrato	MS-DIAL	880	-0.2355	-0.9117
Orgánico	POS	0.71	146.08	C6H11NO3	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	Aminoácido	DNP-Capsicum	7.4294	-0.1919	-0.9194
Orgánico	POS	7.73	308.221	C18H29NO3	Dihydrocapsaicin	Alcaloide	MS-DIAL	854.8	-0.1678	-0.5599
Orgánico	POS	2.04	251.139	C13H18N2O3	N-Cinnamoylputrescine; 3",4"-Dihydroxy	Alcaloide	DNP-Solanaceae	7.7706	-0.0883	-0.8919
Orgánico	POS	7.73	615.435	No record	Unknown597_POS	#N/A	-1	-1	-0.0863	-0.6813
Convencional	POS	0.72	133.06	C4H8N2O3	L-Asparagine	Aminoácido	MS-DIAL	872.18	0.1195	0.6276
Convencional	POS	7.28	137.059	C8H8O2	4'-Hydroxyacetophenone	Aromático	DNP-Solanaceae	6.6464	0.1654	0.8225
Convencional	Neg	0.74	191.013	C6H8O7	Citric acid	Carbohidrato	MS-DIAL	879.552	0.1669	0.8760
Convencional	POS	7.28	611.406	C32H48O6	Azadirachnol	Terpeno	DNP-Capsicum	6.5989	0.2085	0.7410
Convencional	POS	7.28	306.206	C18H27NO3	Capsaicin	Alcaloide	GNPS	0.9	0.3895	0.7921

Tabla 35. Metabolitos Marcadores de *C. baccatum* Orgánico vs Convencional por análisis S-plot

Cosecha	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
Orgánico	POS	0.71	146.08	C6H11NO3	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	Aminoácido	DNP-Capsicum	7.4294	-0.3313	-0.8061
Orgánico	Neg	0.68	215.026	C10H10O4	2,3-Dihydro-5,7-dihydroxy-2-methyl-4H-1-benzopyran-4-one	Benzopirano	DNP-Solanaceae	6.393	-0.3177	-0.9248
Orgánico	Neg	3.25	355.091	C12H20O12	6-O-a-D-Glucopyranuronosyl-D-glucose; a-Pyranose-form	Carbohidrato	DNP-Capsicum	7.694	-0.1344	-0.9064
Orgánico	Neg	3.91	579.116	C26H28O15	NCGC00385770-01!2-(3,4-dihydroxyphenyl)-5-hydroxy-7-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[(3,4,5-trihydroxyoxan-2-yl)oxymethyl]oxan-2-yl]oxychromen-4-one	Flavonoide	GNPS	0.75	-0.1149	-0.5765
Convencional	POS	7.28	137.059	C8H8O2	4'-Hydroxyacetophenone	Aromático	DNP-Solanaceae	6.6464	0.1521	0.7254
Convencional	POS	0.72	133.06	C4H8N2O3	L-Asparagine	Aminoácido	MS-DIAL	872.18	0.1538	0.6829
Convencional	Neg	0.89	191.013	C6H8O7	Citric acid	Carbohidrato	GNPS	1	0.1624	0.4362
Convencional	POS	7.28	306.206	C18H27NO3	Capsaicin	Alcaloide	GNPS	0.9	0.3599	0.7233

Tabla 36. Metabolitos Marcadores de *C.chinense* Orgánico vs Convencional por análisis S-plot

Cosecha	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
Orgánico	POS	0.71	146.08	C6H11NO3	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	Aminoácido	DNP-Capsicum	7.4294	-0.2984	-0.8976
Orgánico	Neg	0.68	215.026	C10H10O4	2,3-Dihydro-5,7-dihydroxy-2-methyl-4H-1-benzopyran-4-one	Benzopirano	DNP-Solanaceae	6.393	-0.1249	-0.6446
Orgánico	POS	0.67	175.117	C6H14N4O2	L-(+)-Arginine	Aminoácido	MS-DIAL	962.75	-0.0920	-0.5478
Orgánico	Neg	9.65	681.347	C35H54O13	4(20),11-Taxadiene-2,5,10,14-tetrol; 14-O-(3S-Hydroxy-2R-methylbutanoyl), 2,5-di-Ac, 10-O-β-D-glucopyranoside	Terpeno	DNP-Solanaceae	7.0926	-0.0877	-0.8643
Convencional	POS	10.6	270.277	#N/A	Unknown262_POS	#N/A	#N/A	#N/A	0.1266	0.8470
Convencional	POS	7.28	137.059	C8H8O2	4'-Hydroxyacetophenone	Aromático	DNP-Solanaceae	6.6464	0.1977	0.7596
Convencional	POS	7.28	611.406	C32H48O6	Azadirachnol	Terpeno	DNP-Capsicum	6.5989	0.2434	0.6973
Convencional	POS	7.28	306.206	C18H27NO3	Capsaicin	Alcaloide	GNPS	0.9	0.4835	0.7778

7.6. Discriminación de accesiones y especies y su huella metabólica

7.6.1. Discriminación de especies de los ajíes de Chincha

El objetivo del presente análisis fue encontrar los biomarcadores que permitan discriminar las especies *C. baccatum*, *C. chinense* y *C. frutescens* de la localidad de Chincha. Para ello, se realizó un análisis Metabolómico non-targeted por UHPLC-HRMS de los extractos metanólicos (80% acuoso) del fruto entero de ajíes liofilizados. Luego de la adquisición de la data, se realizó un análisis multivariado de la data cromatográfica de las 5 accesiones cultivadas en las tres localidades y luego se analizó la data de Chincha. El procedimiento analítico fue el siguiente: Se analizaron extractos etanólicos de 5 accesiones cultivadas en tres localidades (Chincha, Chiclayo y Huaral), lo que dio un total de quince extractos de ajíes con cuatro repeticiones cada una y además utilizando una muestra QC de control calidad. El orden de inyección se observa en la figura 23, en el cual se analizaron 72 muestras (60 extractos, 6QCs, 6Blancos) en los dos modos de ionización ESI (+,-). Esto permitió minimizar en cierta medida los errores aleatorios (los errores aleatorios disminuyen a medida que aumentan el tamaño de la muestra). La data cromatográfica obtenida fue deconvolucionada y alineada con el software MS-DIAL y luego se procedió con la identificación de los metabolitos. Luego se combinaron las tablas de datos de las polaridades ESI (+, -) utilizando el Package de R studio MScombine. Esta etapa permitió correlacionar ambas tablas de datos ESI (+, -) para mantener solo las características comunes y filtrar las pseudo

características en forma de aductos y artefactos de los metabolitos originales. El análisis metabolómico realizado por UHPLC-HRMS es de alto rendimiento capaz de proporcionar cientos de características entre conocidas y desconocidas. Sin embargo, presenta algunas desventajas. Entre las desventajas del análisis se pueden citar las características degeneradas producidas como consecuencia de las interacciones iónicas durante la ionización. Entre las características degeneradas se encuentran los compuestos aductos de los compuestos verdaderos con algunos metales alcalinos presentes en el solvente de extracción, en el sistema del solvente de elución como el agua, acetonitrilo y ácido fórmico, la formación de dímeros o trímeros, isotopos de los metabolitos verdaderos, etc, o posible formación de artefactos producidos durante la elución cromatográfica debido a la probable degradación de la estructura molecular de algunos metabolitos lábiles e inestables. En algunas ocasiones algunos metabolitos puede que no alcancen ser ionizados y quedar fuera del análisis con pérdida de información espectrométrica. El solvente de extracción también juega un rol importante. Los extractos de ajíes liofilizados fueron preparados con un solvente polar (metanol 80%), lo que permite extraer metabolitos polares, y semi-polares, dejando de lado algunos metabolitos apolares como los carotenoides.

Para realizar el filtrado de las características degeneradas fue necesario la utilización de herramientas informáticas en metabolómica computacional. El software MScombine permitió hacer este filtrado.

Posteriormente con la data cromatográfica filtrada se realizó un análisis de componentes principales (PCA). El PCA de esta data combinada se observa en la figura 85. Al analizar el PCA de esta data original con normalización y escalado pareto, se observan 3 cluster de las muestras con tendencia a formar tres categorías de la clasificación por especies. En el PCA, los QC (la muestra de control de calidad constituida por un pull o mezcla de los 60 extractos) son proyectados sobre el centro del eje “y” ($t[0]$), estos a su vez atestiguan la reproducibilidad y robustez del análisis. Las muestras están segregadas a lo largo del eje “y” en función de las especies de ajíes. Del PCA de la figura 81A, se excluyó del análisis las muestras QC y posteriormente se construyó un modelo OPLS-DA (Figura 85B). Del PCA de la figura 85A también se realizó un PCA y OPLS-DA de los perfiles cromatográficos de los extractos de ajíes de la localidad de Chincha, como se muestra en la figura 86. Un análisis supervisado (OPLS-DA) permite forzar la separación de las diferentes clases y obtener un modelo predictivo para la discriminación de especies. Este modelo permite poner en evidencia los biomarcadores asociados a cada clase mediante un análisis S-plot. La calidad de los modelos fue descrita por los valores de R^2 y Q^2 . La bondad de ajuste, R^2 ($0 \leq R^2 \leq 1$) explica la variación, y es una medida cuantitativa de qué tan bien se reproducen matemáticamente los datos del conjunto de entrenamiento. La habilidad predictiva del modelo se evalúa por el Q^2 acumulativo que representa la fracción de la variación de Y que puede ser predecidas por el modelo, y se obtiene de acuerdo con el método de validación cruzada interna del software SIMCA-P 15. En el procedimiento de validación cruzada en SIMCA-P 15, los parámetros se estiman en

una parte de la matriz y la bondad de los parámetros en términos de su éxito se ensayan en la predicción de la otra parte de la matriz. Cada observación se omite una vez durante la validación cruzada. El número de rondas de validación cruzada (por defecto es 7) es el número de grupos, que uno por uno se deja fuera del modelado y se repite durante la validación cruzada. Es decir, el número predeterminado de grupos de validación cruzada es 7 y se excluye una séptima parte de las observaciones durante cada ronda de la validación cruzada. Los valores de R^2 y Q^2 del modelo generados son una medida de la robustez del modelo ($R^2 > 0.8$, $Q^2 > 0.6$).

En los análisis de PCA de las figuras 85 y 86, los errores sistemáticos y aleatorios se observan como variación intraclase. Por tal motivo, la data necesitó ser normalizada. Para ello, se ensayaron 20 métodos de normalización, dichos métodos están localizados en la plataforma del software online NOREVA (57), mostrando el mejor desempeño el método de normalización EigenMS (118,119). El método de normalización EigenMS tiene como objetivo identificar los errores sistemáticos y aleatorios y eliminar sus efectos de la data preservando las diferencias originales de las mismas. Se tuvieron en cuenta los errores aleatorios debido al lugar de cultivo de los ajíes (localidades) y al tipo de práctica agrícola, (en los ajíes de Chíncha se desarrolló una Agricultura Orgánica mientras que con los ajíes de Chiclayo y Huaral se desarrolló una agricultura convencional) y los errores sistemáticos debidos al operador. De esta manera, luego de la normalización por el método EigenMS, se

encontraron 93 errores sistemáticos y aleatorios de complejidad desconocida (características o metabolitos) que fueron eliminados de la data original.

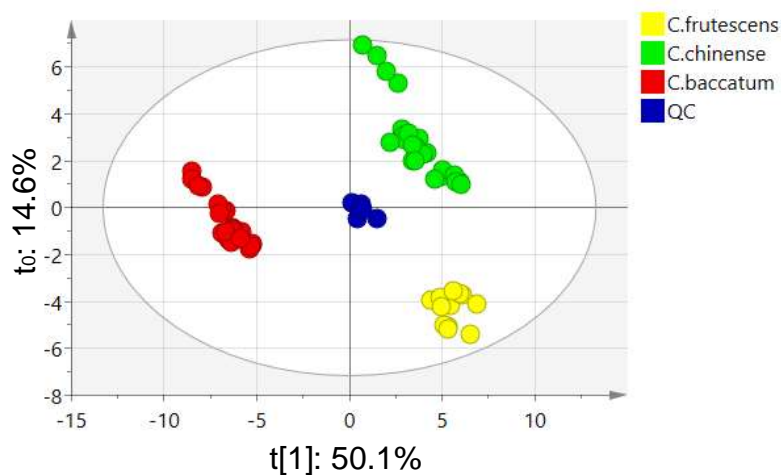
La influencia del método de normalización en la precisión de la clasificación fue evaluada por la curva ROC (Receiver Operating Characteristics) mostrada en la figura 110, con un AUC =1, que significa que es un excelente modelo y que tiene una buena medida de separabilidad.

Posteriormente, la data fue analizada nuevamente por PCA, OPLS-DA (figuras 90 y 91) con el software SIMCA-P 15. De esta data normalizada se extrajo la data de la localidad de Chincha para poder trabajarla separadamente y se analizó por PCA y OPLS-DA y gráficos S-plot. Esta estrategia de análisis permitió mantener la información común de las tres especies en las tres localidades y separar la información de Chincha para el análisis discriminante y análisis S-plot.

El método de normalización EigenMS utilizando el software NOREVA tiene 3 etapas. En una primera etapa del procesamiento de la data se cargó la tabla de datos (subir la data on line en formato *. CSV) en el software NOREVA seleccionando el tipo de data de muestras con QC. La tabla de datos cromatográficos debe ser preparada de tal manera que debe respetar el orden de inyección de las muestras (figura 23). El experimento LCMS se realizó con 72 muestras (60 muestras + 6 QC + 6Blks) de acuerdo con el diseño experimental intercalando cada dos muestras una muestra QC y blanco (Blk). En una segunda etapa, se mantuvo por defecto los

parámetros de procesamiento de la data: 1) criterio de filtrado igual a 0.8, 2) compensación de la varianza de sesgo igual a 0.75, 3) modelo de regresión de ajuste Polinomial y 4) algoritmo de imputación KNN (K-vecinos más cercanos) (120). Esta etapa permitió hacer una corrección basado en el QC (figura 87).

A)



B)

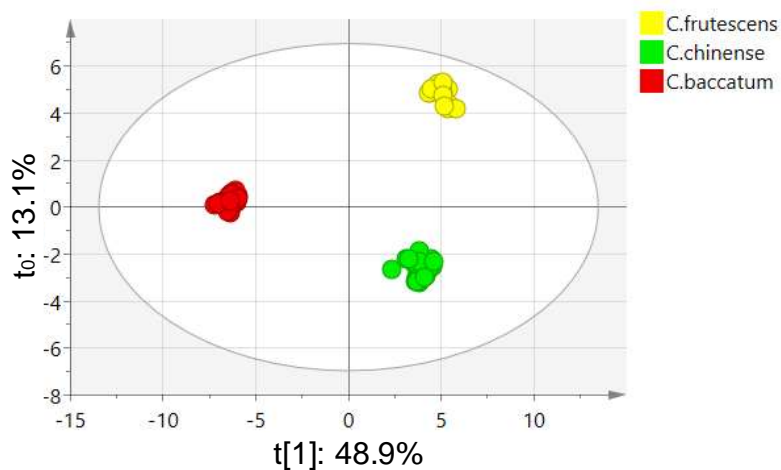
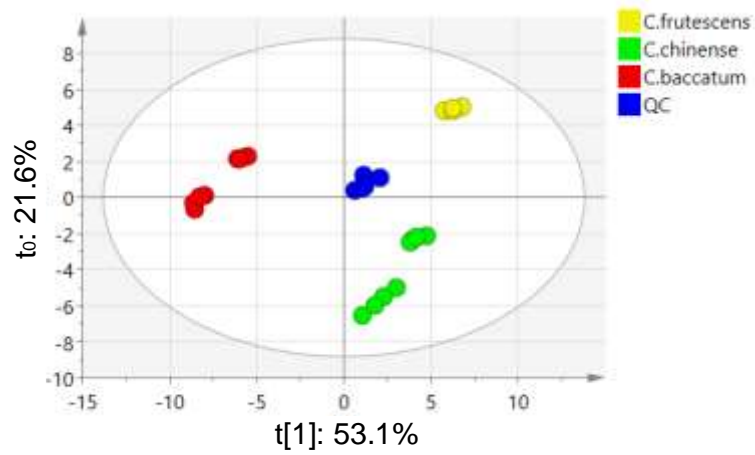


Figura 85. PCA(A) y OPLS-DA (B) de los perfiles UHPLC-HRMS de la localidad de Chincha, Chiclayo y Huaral de la clase especie. El PCA muestra un $R^2X= 0.948$, el OPLS-DA muestra un $R^2Y= 0.990$ y $Q^2 = 0.986$. CV-ANOVA ($p = 0$). La varianza explicada es del 64.7 y 62%, respectivamente.

A)



B)

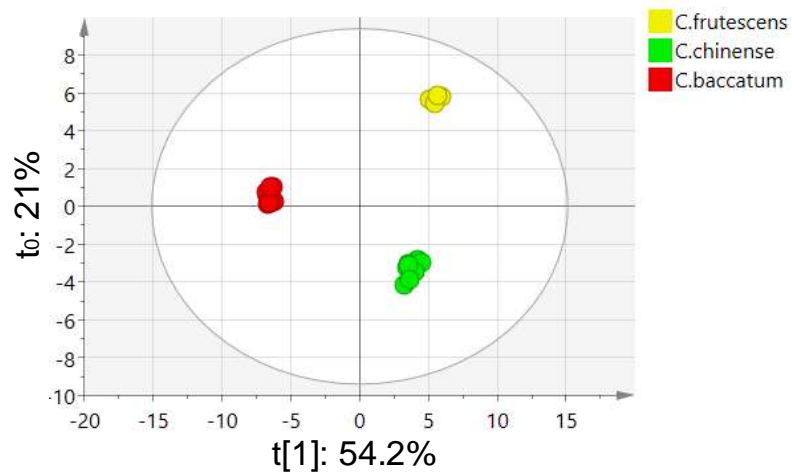


Figura 86. PCA (A) y OPLS-DA (B) de los perfiles UHPLC-HRMS de la localidad de Chíncha de la clase especie antes de ser normalizada con el método EigenMS. El PCA muestra un $R^2X=0.948$, el OPLS-DA muestra un $R^2Y=0.993$ y $Q^2=0.991$. CV-ANOVA ($p=9.86565e-22$). La varianza explicada es del 74.7y 75.2%, respectivamente.

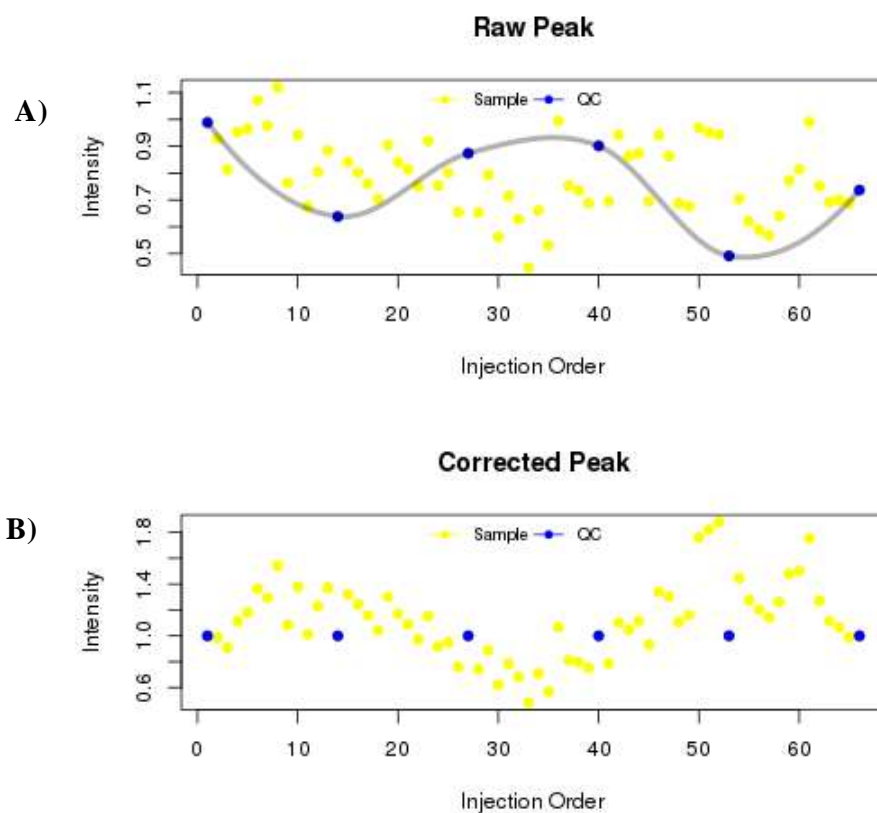


Figura 87. Corrección de picos cromatográficos basados en el QC. En A se observa la data cromatográfica cruda inicial (6 QC). Luego, los picos son corregidos utilizando un ajuste de regresión polinomial. El ajuste permite obtener una intensidad de ajuste del QC constante mostrado en B.

En una tercera etapa, se eligió el método de normalización EigenMS. Luego de aplicar la normalización de la data, los resultados fueron exportados en formato *.CSV y evaluados por PCA, OPLS-DA y analizados por S-plot con el software SIMCA-P 15. La figura 88 muestra los gráficos Boxplots de la distribución de la data cromatográfica antes y después de la normalización de la data por el método EigenMS.

El software online NOREVA permitió hacer una corrección de deriva de señal utilizando un enfoque denominado corrección robusta de señal LOESS (suavizado

de diagrama de dispersión estimada localmente) basado en las muestras de control de calidad (QC-RLSC). La deriva en la señal para cada característica metabólica en cada muestra sujeto se corrige observando el cambio en la señal para la misma característica metabólica en las muestras QC y corrigiendo en las muestras sujeto el cambio temporal en la señal relacionada con lo observado para las muestras QC.

La deriva de la intensidad de la señal a lo largo del tiempo es un factor de confusión importante en los estudios metabólicos. Por este motivo es importante el análisis periódico de las muestras QC junto con las muestras sujeto. Al final de la ejecución experimental, y después de la deconvolución cromatográfica, cada característica metabólica detectada se normaliza a la muestra de QC utilizando el algoritmo QC-RLSC. Se ajusta una función de suavizado no lineal estimada localmente de bajo orden (LOESS) a los datos del QC con respecto al orden de inyección. Luego se interpola una curva de corrección para toda la ejecución analítica, a la cual se normaliza el conjunto de datos totales para esa característica. Con este procedimiento, se minimiza cualquier atenuación de la respuesta máxima durante una ejecución analítica (es decir, cualquier factor de confusión debido al orden de inyección) (57,121).

Para evaluar el desempeño del método de normalización en la reducción de la variación intraclase, el software NOREVA muestra los resultados del PCA. Los resultados del PCA permitieron evaluar el éxito de la normalización. En la figura 89B se observa el PCA luego de la normalización por el método EigenMS con

variación intraclase nula, donde C. bac = *C. baccatum*, C. chin = *C. chinense* y C. frut = *C. frutescens*.

En la figura 90 se muestra el PCA y OPLS-DA de la data Normalizada (tres localidades) por el método EigenMS utilizando el software SIMCA-P 15. La varianza explicada es de 65.4% para ambos modelos, respectivamente. De este PCA, se extrajo la data de la localidad de Chíncha para realizar el PCA y OPLS-DA que se muestra en la figura 91.

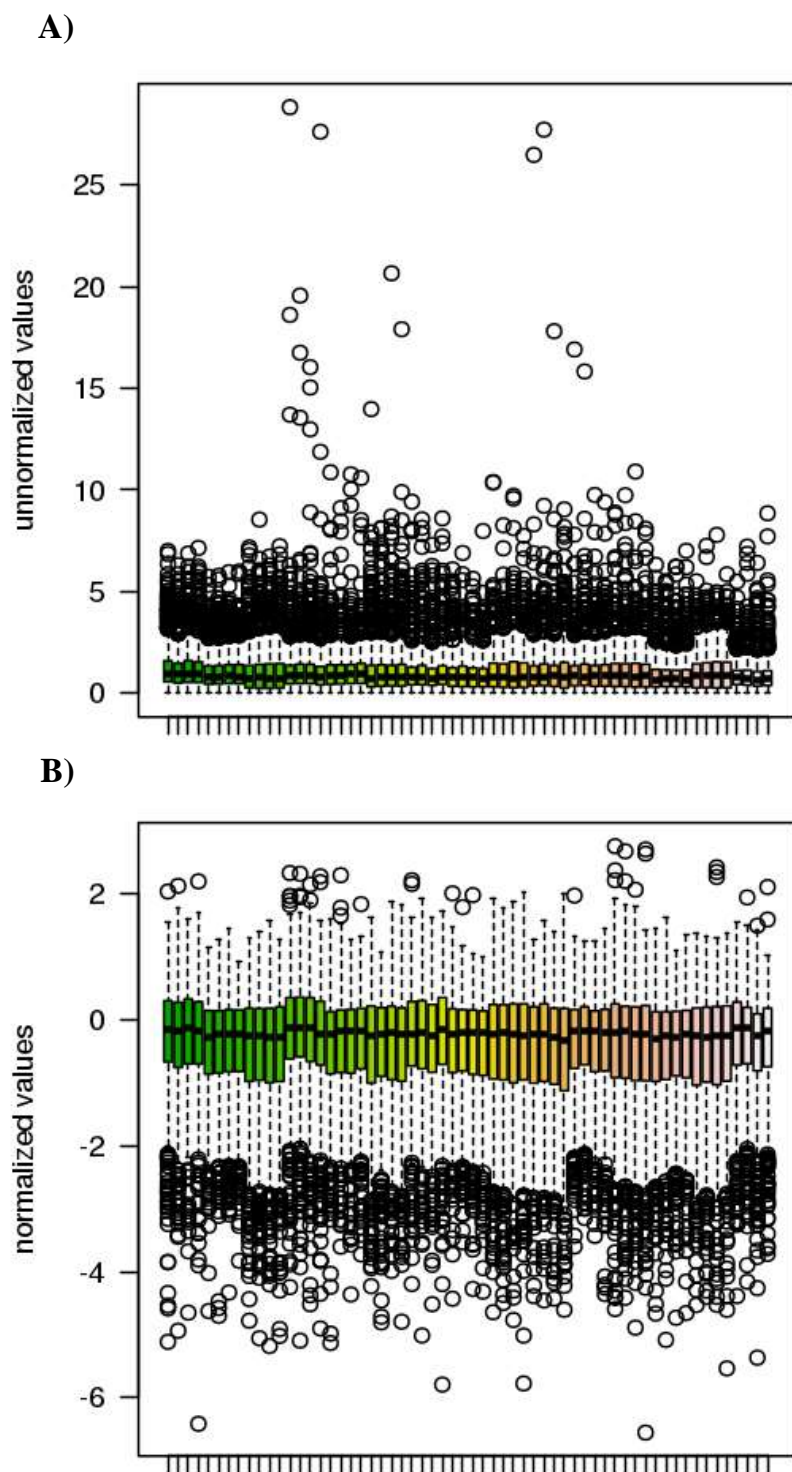


Figura 88. Diagramas Boxplots de la data cromatográfica utilizando el método de normalización EigenMS. En A se observa el diagrama boxplots de intensidades de la data cromatográfica antes de la normalización, en B se observan los diagramas boxplots de intensidades de la data normalizada, no se muestran las intensidades de los QCs. Cada intensidad aparece en el orden de ejecución de la corrida cromatográfica.

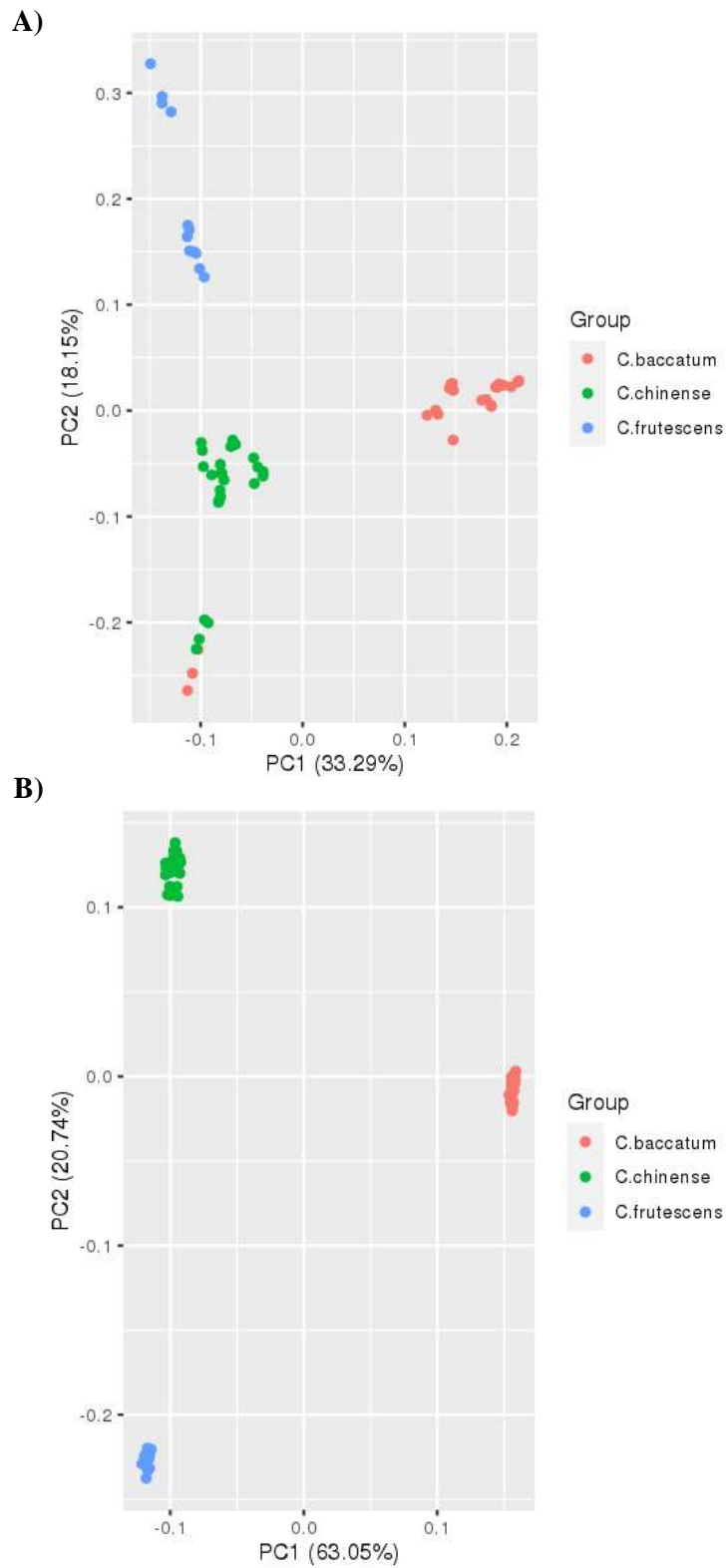


Figura 89. PCA para evaluar el desempeño del método de normalización EigenMS. Las figuras 36A y 36B muestran el PCA antes y después de la normalización por EigenMS. Se observa un buen desempeño con un aumento de la varianza explicada desde el 51.44% hasta un 83.52%.

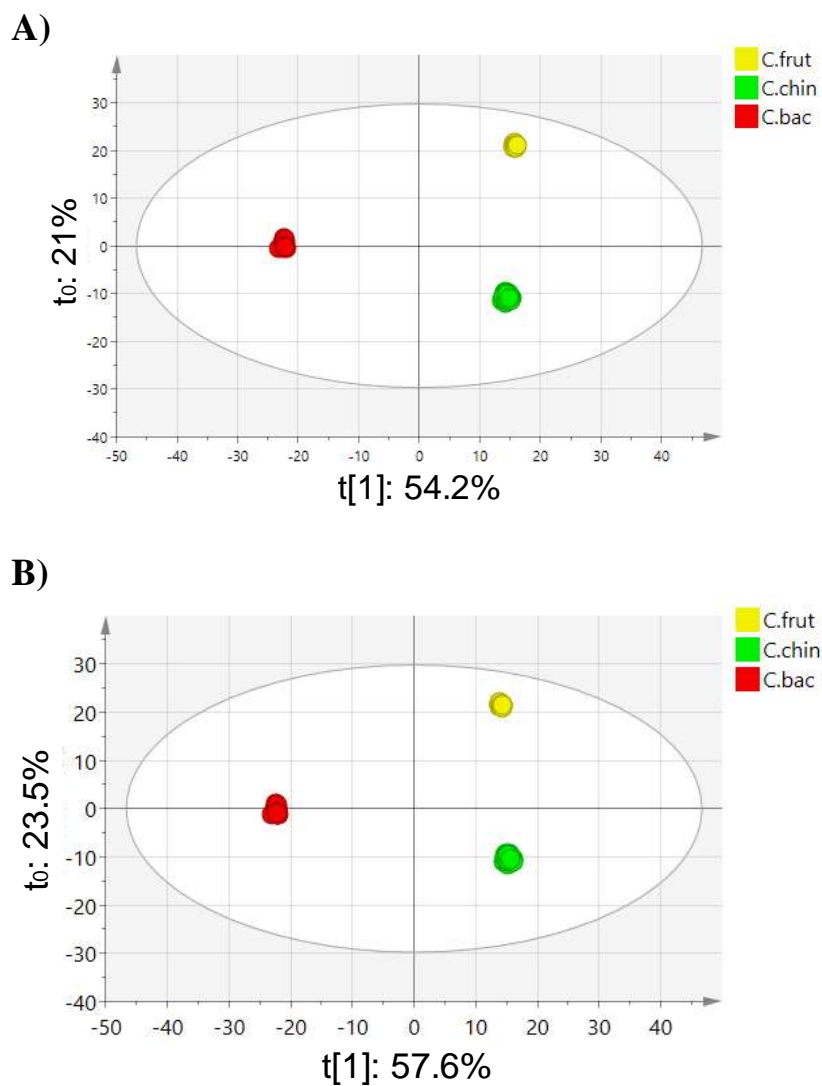


Figura 90. PCA (A) y OPLS-DA(B) de la data cromatográfica de los ajíes de las tres localidades Chincha, Chiclayo y Huaral, normalizada por el método EigenMS de la clase especie. El PCA muestra un $R^2X= 0.877$, el OPLS-DA muestra un $R^2Y= 0.999$; $Q^2 = 0.999$. CV-ANOVA ($p = 0$). La varianza explicada es del 81.1% para ambos modelos.

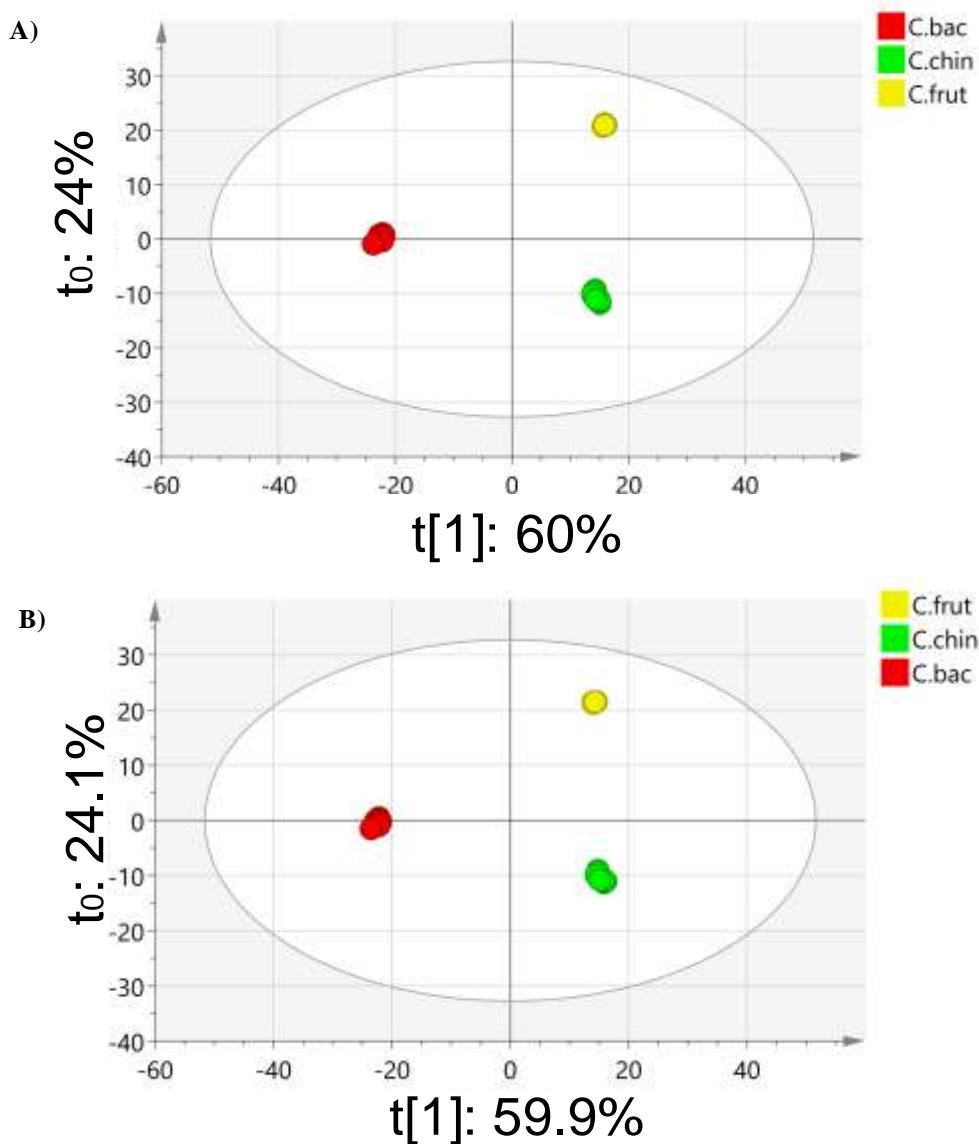


Figura 91. PCA, data cromatográfica de los ajíes de la localidad de Chincha normalizada por el método EigenMS. El PCA muestra un $R^2X = 0.840$, el OPLS-DA muestra un $R^2Y = 0.995$ y $Q^2 = 0.765$. CV-ANOVA ($p = 0.000632812$). La varianza explicada es del 84% para ambos modelos.

Para mostrar los metabolitos discriminantes entre las especies se realizó un OPLS-DA enfrentando a 2 individuos uno a uno. En la figura 92 y 93 se muestra el OPLS-DA y gráfico S-plot de las clases *C. frutescens* vs *C. chinense*, *C. baccatum* vs *C. chinense* y *C. baccatum* vs *C. frutescens* de la localidad de Chincha para la identificación de biomarcadores.

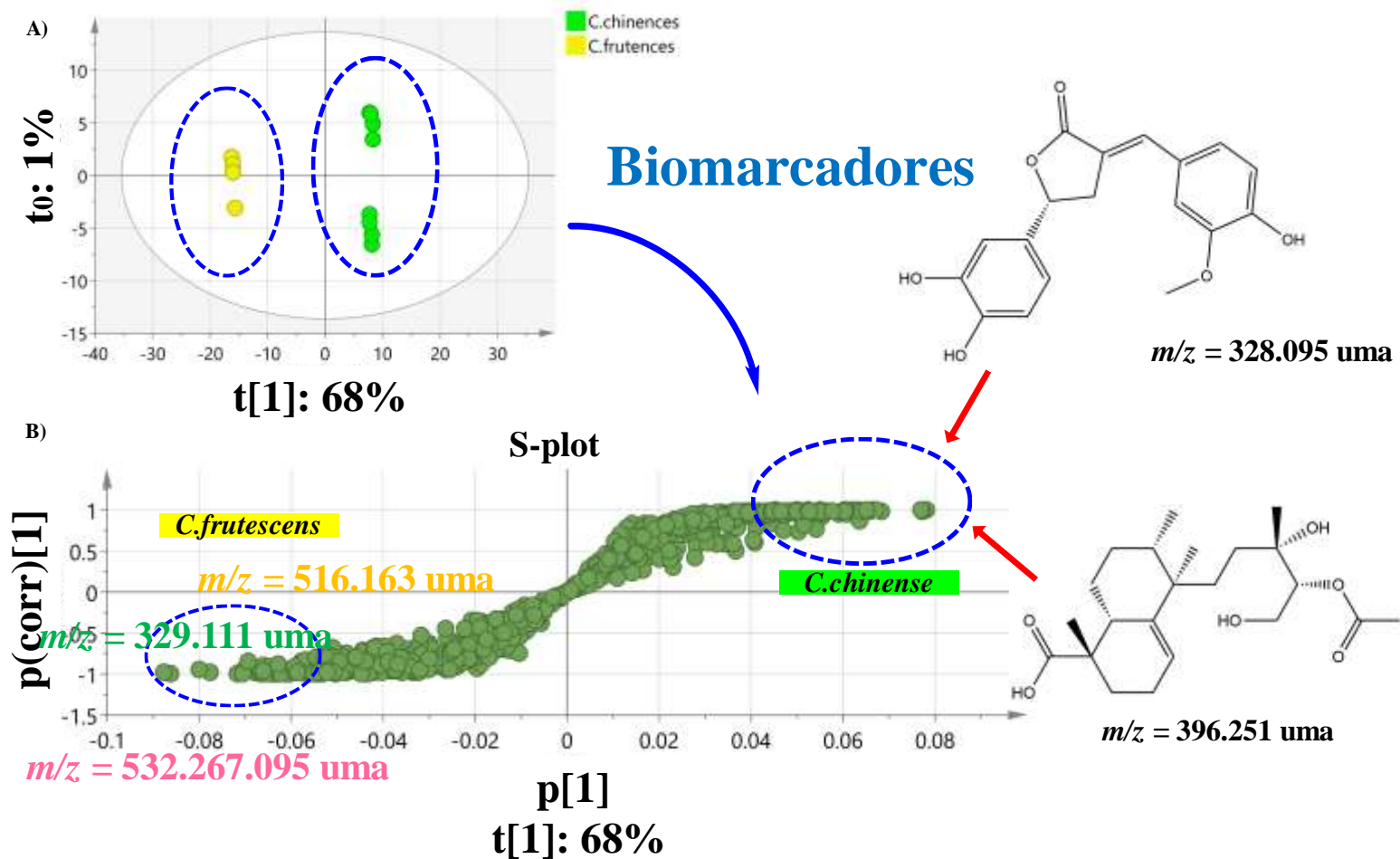


Figura 92. Identificación de biomarcadores mediante análisis discriminante OPLS-DA (A) y S-plot (B) de las especies *C. frutescens* y *C.chinense*. Los puntos extremos del diagrama S-plot reflejan los principales Biomarcadores. En la tabla 37 se muestran los biomarcadores identificados por este método que son únicos para cada especie.

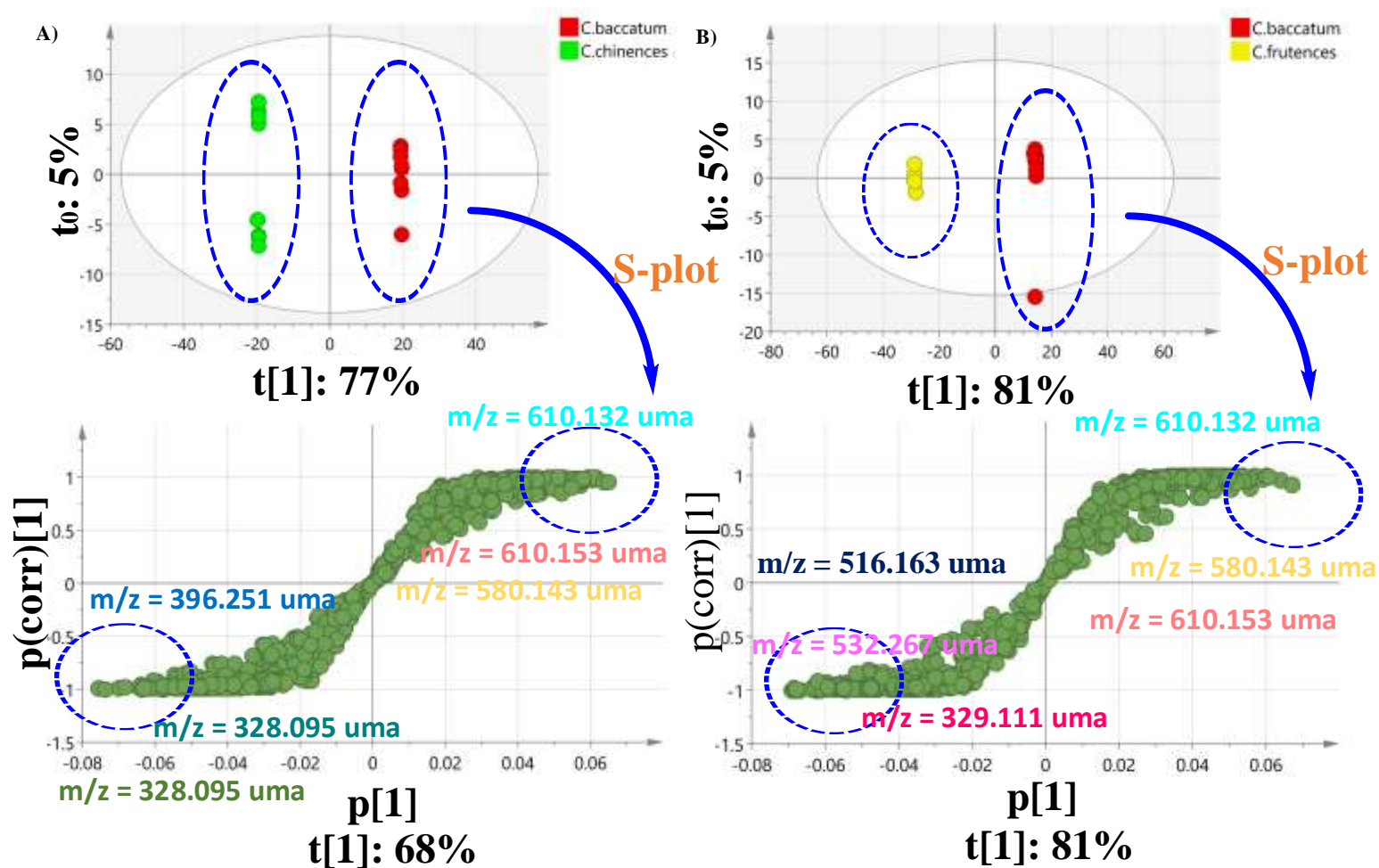


Figura 93. Identificación de biomarcadores mediante análisis discriminante OPLS-DA (A) y S-plot (B) de las especies *C. baccatum* vs *C. chinense* y *C. baccatum* vs *C. frutescens*. Los puntos extremos del diagrama S-plot reflejan los principales Biomarcadores, se indican la relación m/z de los biomarcadores identificados. En la tabla 37 y 38 se muestran los biomarcadores identificados por este método que son únicos para cada especie

Un análisis S-plot permite la identificación de los biomarcadores (figura 92B). Esta combina las variables de contribución (covarianza) o loadings “p[1]” y las variables de confianza o correlación “p(corr)[1]” del modelo OPLS-DA. Los puntos extremos representan las variables más discriminantes (biomarcadores). Los biomarcadores ideales tienen una gran contribución y una alta confiabilidad con pequeño riesgo de correlaciones espurias. Los picos con baja magnitud/intensidad están cercanos al nivel del ruido con un alto riesgo de correlaciones espurias (puntos cercanos al origen). En la figura 92B los puntos extremos comunican los biomarcadores con una alta confiabilidad. En las figuras 93A y 93B se observa el OPLS-DA y diagrama S-plot para las clases especies *C. baccatum* vs *C. chinense* y *C. baccatum* vs *C. frutescens*.

En las tablas 37 y 38 se reúnen las hipótesis de identificación *in Silico* e identificaciones con espectros experimentales de sustancias naturales conocidas de la data cromatográfica de los ajíes de Chincha. Para la identificación *in Silico* se consultaron las bases de datos del Diccionario de Productos Naturales de acuerdo al género *Capsicum* y a la familia Solanaceae. También se consultaron las bases de datos interna del software MSFinder de sustancias naturales (KNAPSAcK, Pubchem, PlantCyc, DrugBank, CheBI, NNPDB, UNDP, LipidMAPS). Para la identificación con espectros experimentales se utilizaron las bases de datos del software MS-DIAL (All Public DBs) y para la identificación con la base de datos GNPS se utilizó el software Metgem. En la identificación *in Silico*, una identificación en función del género *Capsicum* o la familia Solanaceae tuvo mayor

peso en comparación con la identificación con bases de datos de sustancias naturales. Estas bases de datos son privilegiadas. En la columna de Score, el valor (x/1000) (Dot product), refleja la similaridad entre el espectro experimental y el espectro experimental de la base de datos interna. El valor (x/1) refleja la similaridad entre un espectro experimental y la base de datos de GNPS. El valor Total score (x/10) refleja la similaridad entre un espectro experimental y un espectro *in Silico*. Valores de Dot product > 750 fueron considerados como una buena identificación (MS-DIAL). Valores de Total Score > 6 fueron considerados como una buena identificación (*in Silico*). Valores con score > 0.7 fueron considerados como una buena identificación (para GNPS).

Tabla 37. Resultados del análisis S-plot, biomarcadores a partir del modelo OPLS-DA de las especies *C. frutescens* (*C. frut*) y *C. chinense* (*C. chin*).

Especie	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
<i>C.frut</i>	POS	2.54	516.163	C26H28O11	8-C-β-D-Glucopyranosyl-4',5,7-trihydroxyflavone; 3"-O-(2-Methylbutanoyl)	Flavonoide	DNP-Solanaceae	6.9792	-0.0668	-0.9418
<i>C.frut</i>	POS	2.57	329.111	C14H19NO8	4-(2-Nitroethyl)phenol; O-β-D-Glucopyranoside	Aromático	DNP-Capsicum	7.6203	-0.0619	-0.9523
<i>C.frut</i>	Neg	4.18	532.267	C29H40O9	3-Hydroxy-19-oxocarda-14,20(22)-dienolide; 3-O-β-D-Glucopyranoside	Esteroides	DNP-Solanaceae	7.2559	-0.0624	-0.9584
<i>C.frut</i>	POS	1.78	224.129	C12H18NO3+	4-Hydroxybenzoylcholine	Alcaloide	MS-DIAL	880	-0.0604	-0.9920
<i>C.frut</i>	Neg	3.28	283.074	C10H13N5O3S	5'-Thioadenosine	Nucleósido	DNP-Solanaceae	7.2346	-0.0662	-0.9565
<i>C.frut</i>	POS	3.85	432.107	C21H20O10	Isovitexin	Flavonoide	GNPS	0.9	-0.0640	-0.9942
<i>C.frut</i>	POS	2.57	329.111	C14H19NO8	4-(2-Nitroethyl)phenol; O-β-D-Glucopyranoside	Aromático	DNP-Capsicum	7.6203	-0.0619	-0.9523
<i>C.frut</i>	POS	7.2	275.189	C17H25NO2	p-Menthan-3-ol; 2-Aminobenzoyl	Terpenoide	DNP-Solanaceae	7.211	-0.0406	-0.9015
<i>C.chin</i>	Neg	2.85	444.178	C24H28O8	8,9-Dihydroxy-3,10(14),11(13)-guaiatrien-12,6-olide; 8-O-(2-acetoxymethyl-2E-butenoyl), 9-Ac	Terpenoide	DNP-Solanaceae	6.6504	0.0770	0.9942
<i>C.chin</i>	Neg	9.23	706.356	C37H54O13	2,3,16-Trihydroxy-28-nor-15-oxo-12-oleanen-23-oic acid; 3-(4-O-Acetyl-β-D-glucuronopyranoside)	Terpenoide	DNP-Capsicum	6.9143	0.0443	0.9926
<i>C.chin</i>	Neg	9.23	774.340	C47H50O10	4-O-a-D-Xylopyranosyl-D-xylose; Benzyl glycoside, 3'-Ac, tetrabenzyl	Carbohidrato	DNP-Solanaceae	6.0649	0.0452	0.9932
<i>C.chin</i>	Neg	8.43	624.315	C32H48O12	2,4,8,9,11,12,17-Heptahydroxy-5,13-briaradien-18,7-olide; 4-Octanoyl, 2,9-di-Ac	Terpenoide	DNP-Solanaceae	7.5239	0.0527	0.9883
<i>C.chin</i>	Neg	9.91	736.367	C38H56O14	8,14-Epoxy-3,4-dihydroxycardanolide; 20,22-Didehydro, 4-Ac, 3-O-[β-D-glucopyranosyl-(1→4)-α-D-oleandropyranoside]	Esteroides	DNP-Solanaceae	7.2105	0.0772	0.9773
<i>C.chin</i>	POS	9.22	396.251	C22H36O6	13,14,15-Trihydroxy-1(10)-halimen-18-oic acid; 14-Ac	Terpenoide	DNP-Solanaceae	7.3379	0.0446	0.9783
<i>C.chin</i>	POS	9.22	682.356	C35H54O13	3,8,12,14,17-Pentahydroxypregn-5-en-20-one; 12-Ac, 3-O-[β-D-digitoxopyranosyl-(1→4)-β-D-digitoxopyranoside]	Esteroides	DNP-Capsicum	6.8605	0.0440	0.9796
<i>C.chin</i>	Neg	1.05	328.095	C18H16O6	3,3',4,4'-Tetrahydroxy-9-nor-8,8'-lign-7'-en-9',7'-olide; 3'-Me ether	Lignano	DNP-Solanaceae	7.4113	0.0435	0.9860

Tabla 38. Resultados del análisis S-plot, biomarcadores a partir del modelo OPLS-DA de la especie *C. baccatum* (C.bac)

Especie	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
<i>C.bac</i>	POS	3.36	458.164	C17H30O14	α -L-Rhamnopyranosyl-(1 \rightarrow 2)-[β -D-glucopyranosyl-(1 \rightarrow 4)]-L-arabinose	Carbohidrato	DNP-Capsicum	6.6133	0.0552	0.9395
<i>C.bac</i>	POS	4.21	666.143	C29H30O18	Kaempferol 3-glycosides; 3-O-[β -D-Apiofuranosyl-(1 \rightarrow 3)-6-O-malonyl- β -D-glucopyranoside]	Flavonoide	DNP-Solanaceae	6.9937	0.0414	0.8964
<i>C.bac</i>	Neg	1.23	412.079	C21H16O9	Artemininorin B	Cumarina	DNP-Solanaceae	6.9416	0.0410	0.9481
<i>C.bac</i>	Neg	4.63	194.058	C10H10O4	3-(2,3-Dihydroxyphenyl)-2-propenoic acid; 3-Me ether	Aromático	DNP-Solanaceae	7.5411	0.0427	0.9563
<i>C.bac</i>	POS	3.91	580.143	C26H28O15	Leucoside	Flavonoide	GNPS	0.91	0.0536	0.9988
<i>C.bac</i>	Neg	3.91	580.143	C26H28O15	NCGC00385770-01!2-(3,4-dihydroxyphenyl)-5-hydroxy-7-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[(3,4,5-trihydroxyoxan-2-yl)oxymethyl]oxan-2-yl]oxychromen-4-one	Flavonoide	GNPS	0.75	0.0569	0.9976
<i>C.bac</i>	Neg	3.81	610.132	C30H26O14	Quercetin 3-glycosides; 3-O-[4-Hydroxy-E-cinnamoyl-(\rightarrow 3)- β -D-glucopyranoside]	Flavonoide	DNP-Solanaceae	7.8299	0.0593	0.9957
<i>C.bac</i>	POS	3.8	610.153	C27H30O16	Rutin	Flavonoide	GNPS	0.96	0.0565	0.9788

En la figura 94 se observa el perfil metabolómico en el diagrama Heatmap de la clase especie de los ajíes de la localidad de Chincha construido a partir del modelo PLS-DA de las 824 características VIP (variables de importancia). La imagen coloreada comunica claras diferencias entre las 3 especies de ajíes *C. baccatum*, *C. chinense* y *C. frutescens*, de acuerdo al agrupamiento jerárquico (HCA) realizado tanto para filas y columnas en la matriz de datos. En las filas se observa el agrupamiento jerárquico por especies, y en columnas el agrupamiento jerárquico de los metabolitos. Las observaciones han sido ordenadas de acuerdo a su similitud. La escala de colores muestra los valores altos y bajos con valores similares uno cerca del otro. Como distancia de medida se utilizó la distancia Euclidiana y como algoritmo de agrupamiento se utilizó el algoritmo Ward.

En la figura 95 se observa los diagramas Boxplots de las áreas normalizadas para los cuatro metabolitos mayoritarios identificados en las muestras de ajíes de la localidad de Chincha. Se observa que la especie *C. frutescens* posee el mayor contenido de capsaicina y dihidrocapsaicina mientras que la especie *C. baccatum* posee el mayor contenido de ácido cítrico. Se puede considerar el ácido cítrico como un biomarcador de esta especie. En las figuras 96-98 se observan los diagramas Boxplots de los principales metabolitos biomarcadores identificados por análisis discriminante OPLS-DA y gráficos S-plot para las tres especies *C. baccatum*, *C. chinense* y *C. frutescens*. De acuerdo con los diagramas boxplots, estos metabolitos biomarcadores son específicos de las especies estudiadas y en general por cada uno de los perfiles Metabolómicos individuales para cada especie.

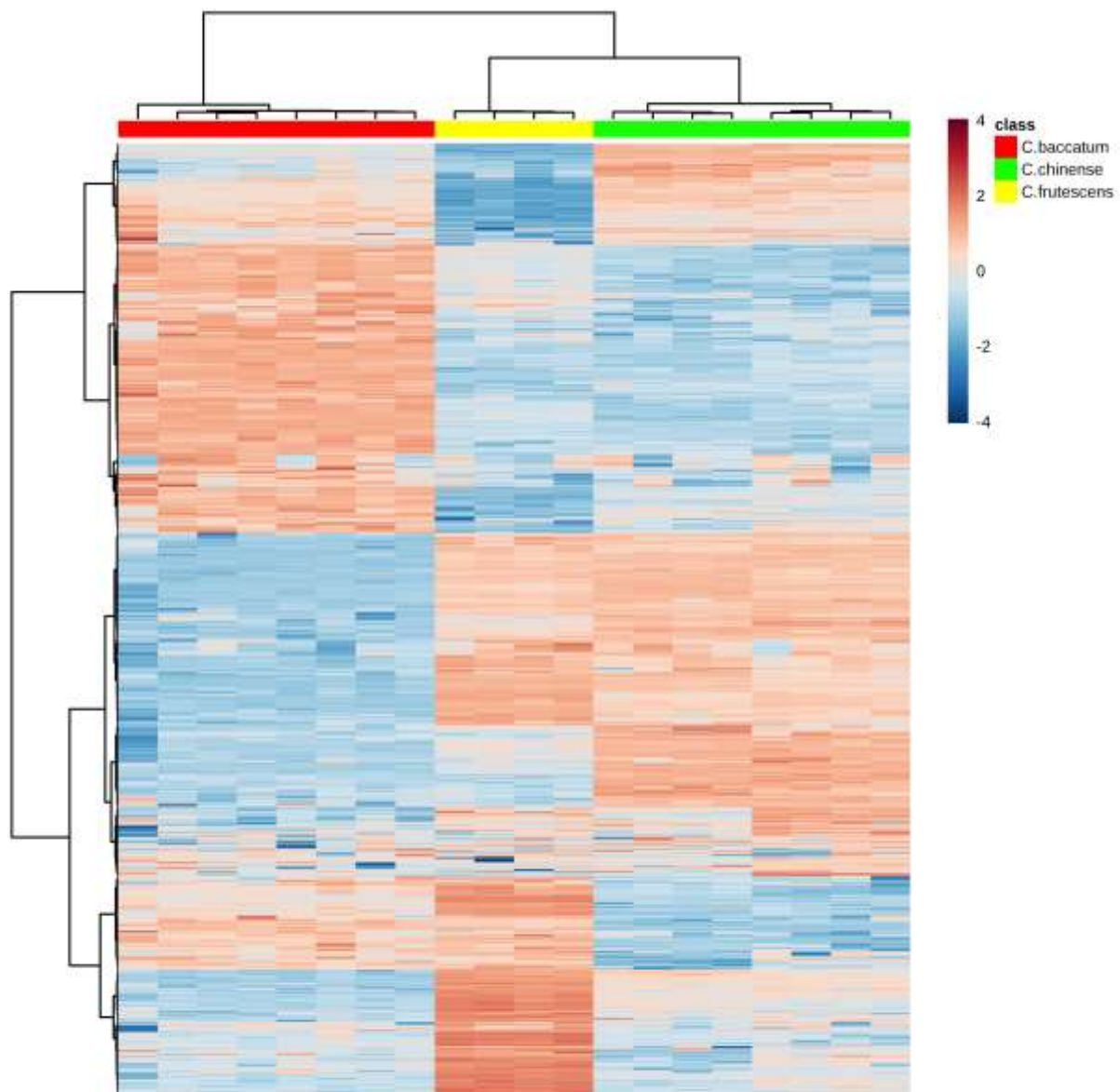


Figura 94. Heatmap de las 824 características de metabolitos no volátiles a partir del modelo PLS-DA de la clase especie del fruto entero de ají de la localidad de Chincha.

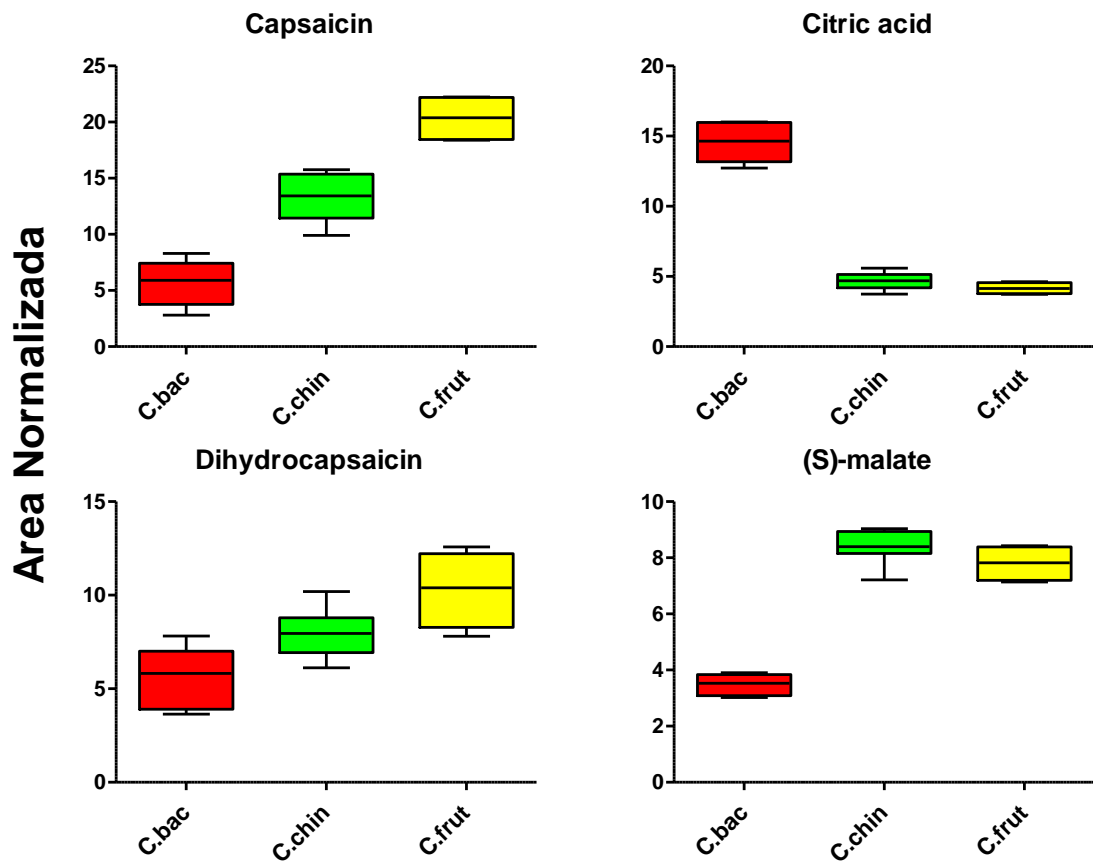


Figura 95. Diagramas Boxplots de Metabolitos con mayores áreas relativas en las 3 especies de ajíes de la localidad de Chincha; C.bac = *C. baccatum*, C. chi = *C. chinense*, C. frut = *C. frutescens*

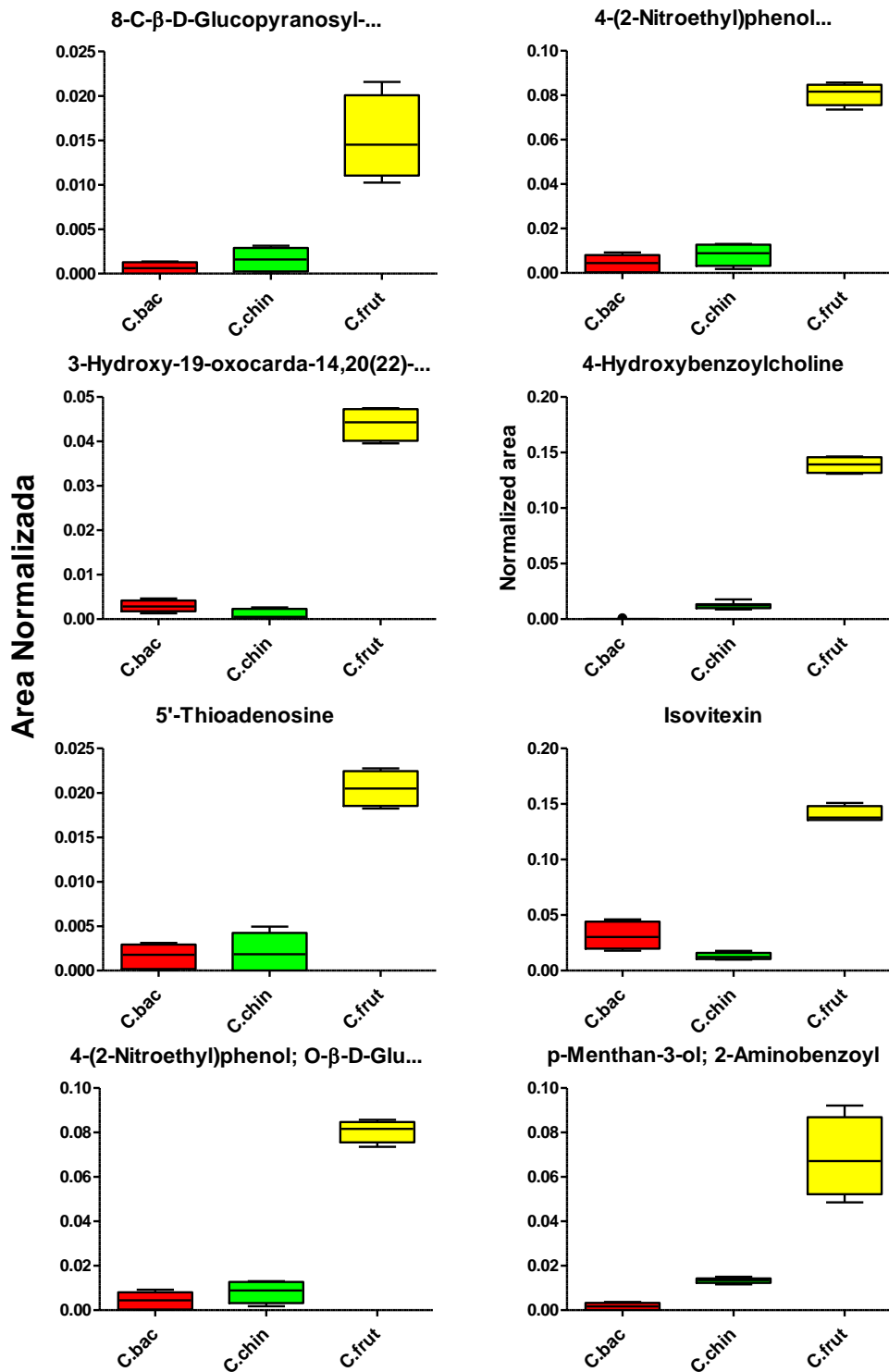


Figura 96. Boxplots de las áreas metabolitos biomarcadores de la especie *C. frutescens*

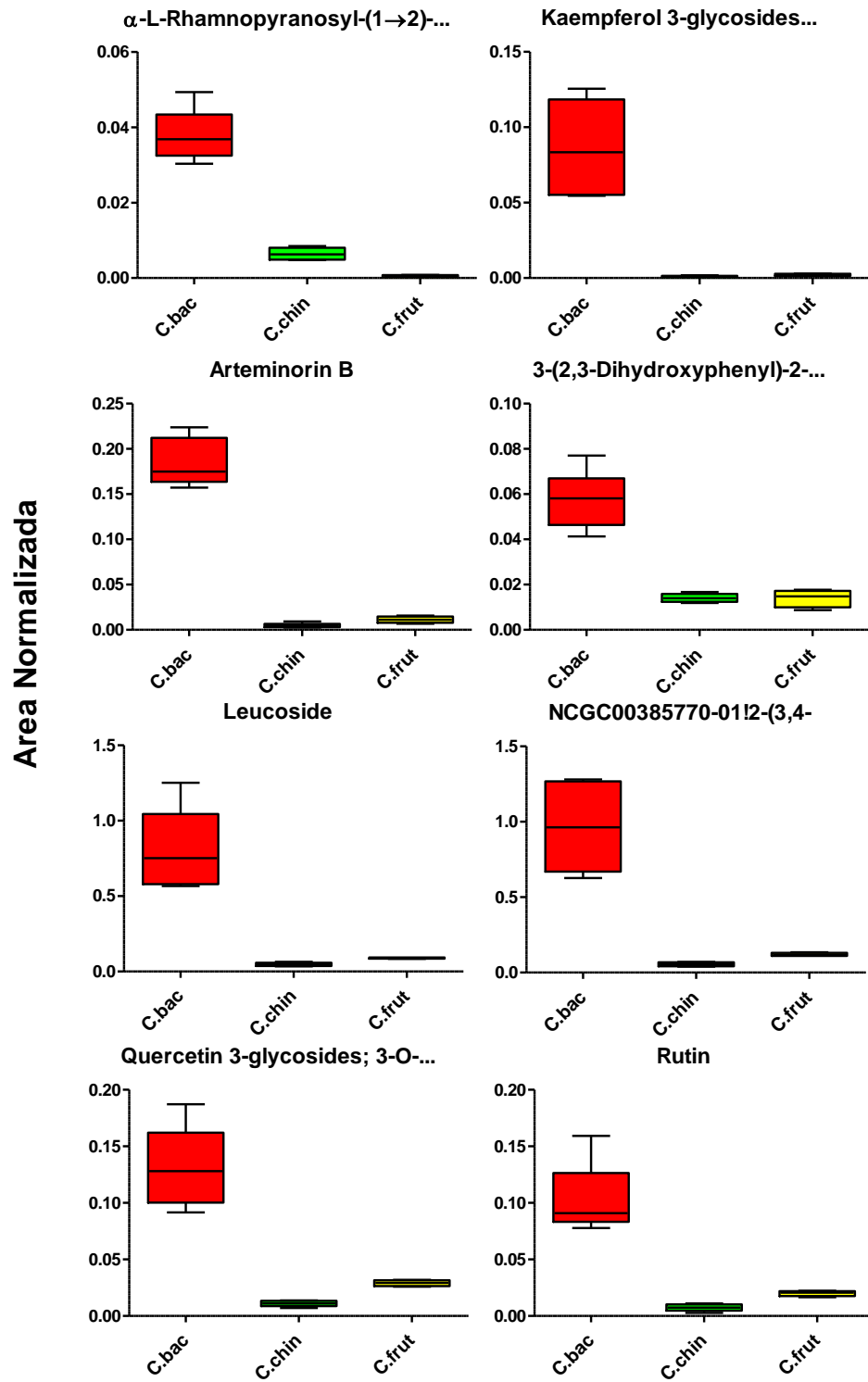


Figura 97. Boxplots de los metabolitos biomarcadores de la especie *C. baccatum*

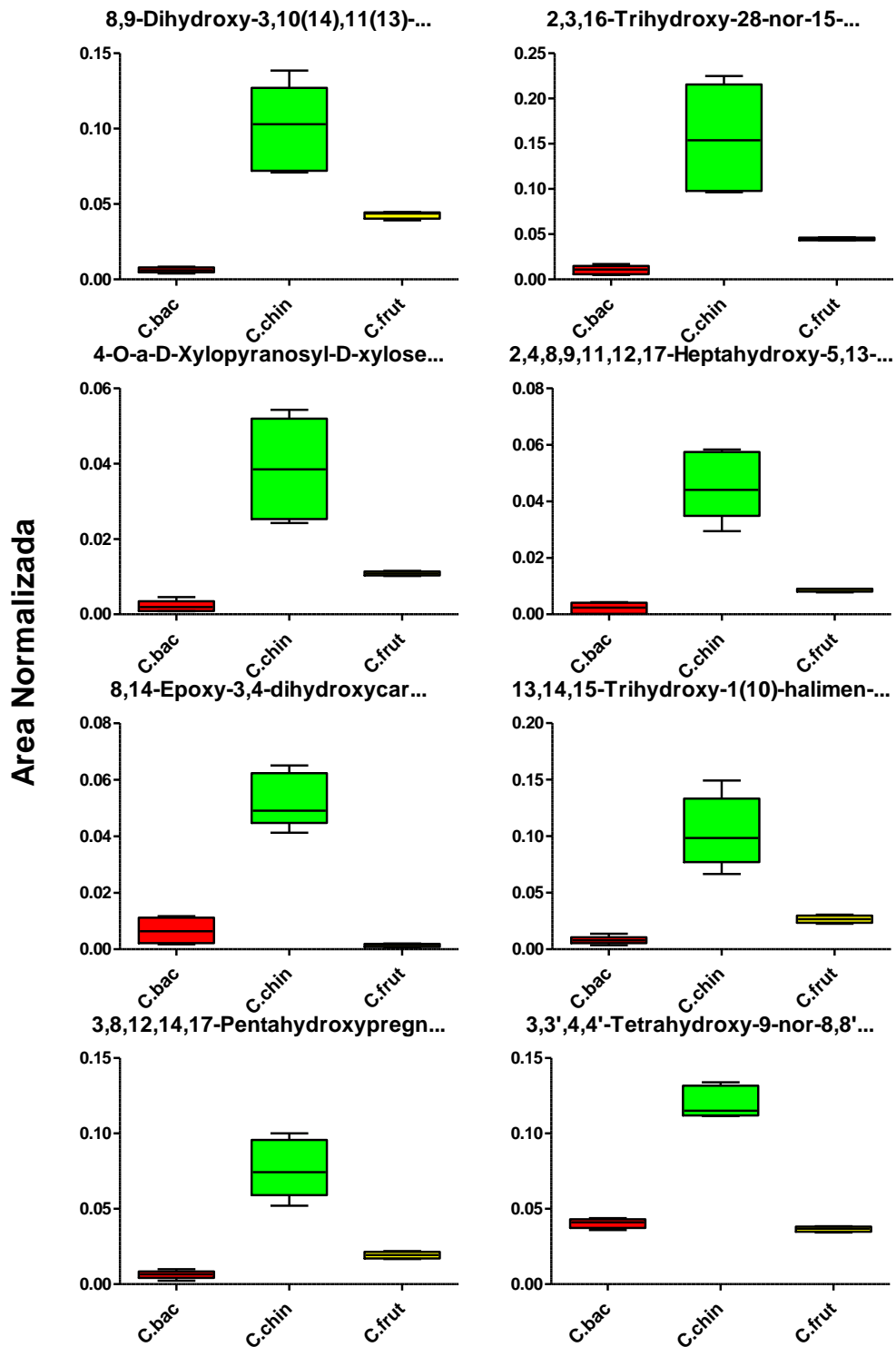


Figura 98. Boxplots de los metabolitos biomarcadores de la especie *C. chinense*

7.6.2. Discriminación de accesiones de los ajíes de Chincha

Para la discriminación de las accesiones de ajíes de la localidad de Chincha, se evaluaron el PCA y OPLS-DA (figura 99) de los perfiles cromatográficos obtenido por UHPLC-HRMS de los ajíes de la localidad de Chincha, Chiclayo y Huaral. Antes de realizar el PCA de esta data cromatográfica, se combinaron las tablas de datos de las polaridades ESI (+,-) utilizando el Package de R studio MScombine. Esta etapa permitió correlacionar ambas tablas de datos ESI (+,-) para mantener solo las características comunes y filtrar las pseudo características en forma de aductos y artefactos de los metabolitos originales.

Del PCA de la figura 99A, se excluyeron los QC y posteriormente se construyó un modelo OPLS-DA (Figura 99B). Un análisis supervisado (OPLS-DA) permitió forzar la separación de las diferentes clases y obtener un modelo predictivo para la discriminación de accesiones.

Del PCA de la figura 99A, se excluyó la data cromatográfica de la localidad de Chincha para construir con esta data un modelo PCA y OPLS-DA como se muestra en la figura 100.

Del PCA y OPLS-DA de las figuras 99A y 100A se observa la variación intracalse de las accesiones. Como en el caso del análisis por especies, la variación proviene de los errores aleatorios y sistemáticos. Para eliminar esta variación, se aplicó el

método de normalización Eigen-MS (118), utilizando el software online NOREVA (57). En una primera etapa del procesamiento de la data se cargó la data cromatográfica en el software NOREVA seleccionando el tipo de data a ingresar (muestras con QC). Al ingresar la data cromatográfica se respetó el orden de inyección de las muestras (figura 23). En una segunda etapa, se mantuvo por defecto los parámetros de procesamiento de la data: 1) criterio de filtrado igual a 0.8; 2) compensación de varianza de sesgo igual a 0.75; 3) modelo de regresión de ajuste Polinomial y 4) algoritmo de imputación KNN (K-vecinos más cercanos) (120). Esta etapa permitió hacer una corrección basada en las muestras QC (figura 48). En una tercera etapa, se eligió el método de normalización EigenMS. De esta manera, luego de la normalización por el método EigenMS, se encontraron 132 errores aleatorios y sistemáticos de complejidad desconocida (características o metabolitos) que fueron eliminados de la data original. La figura 102 muestra los diagramas boxplots de la distribución de la data cromatográfica antes y después de la normalización por el método EigenMS. Para evaluar el desempeño del método de normalización, el software NOREVA muestra los resultados del PCA de la data normalizada (figura 103B).

En la figura 104 se muestra el PCA y OPLS-DA de la data normalizada por el método EigenMS utilizando el software SIMCA-P, la varianza explicada es de 65.4% para ambos modelos, respectivamente. De estos modelos se extrajo la información de Chincha para construir los modelos de PCA y OPLS-DA de esta localidad, como se muestra en la figura 105. En las tablas 39 y 40 se reúnen las

hipótesis de identificación *in Silico* y las identificaciones con espectros experimentales de sustancias naturales conocidas. Para la identificación *in Silico* se consultaron las bases de datos del Diccionario de Productos Naturales de acuerdo al género *Capsicum* y a la familia Solanaceae. También se consultaron las bases de datos de sustancias naturales utilizando el software MSFinder (KNAPSAcK, Pubchem, PlantCyc, DrugBank, CheBI, NANPDB, UNDP, LipidMAPS). Para la identificación con espectros experimentales se utilizaron las bases de datos del software MS-DIAL (All Public DBs) y la base de datos GNPS con la ayuda del software Metgem. En la identificación *in Silico*, una identificación en función del género *Capsicum* o la familia Solanaceae tuvo mayor peso en comparación con la identificación con bases de datos de sustancias naturales. Estas bases de datos son privilegiadas. En la columna de score, el valor (x/1000) (Dot product), refleja la similaridad entre el espectro experimental y el espectro experimental de la base de datos interna. El valor (x/1) refleja la similaridad entre un espectro experimental y la base de datos de GNPS. El valor Total score (x/10) refleja la similaridad entre un espectro experimental y un espectro en Silico. Valores de Dot product > 750 fueron considerados como una buena identificación (All public de MS-DIAL). Valores de Total Score > 6 fueron considerados como una buena identificación (*in Silico*). Valores con score > 0.7 fueron considerados como una buena identificación (para GNPS).

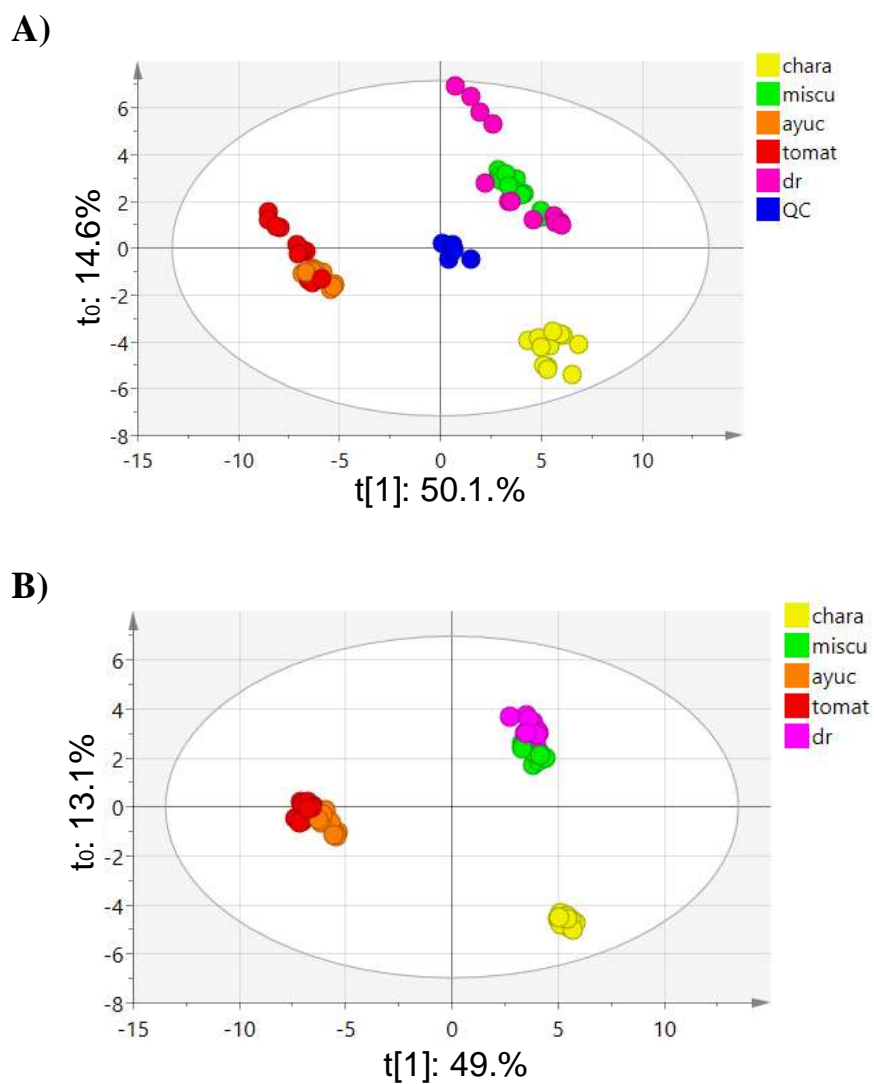
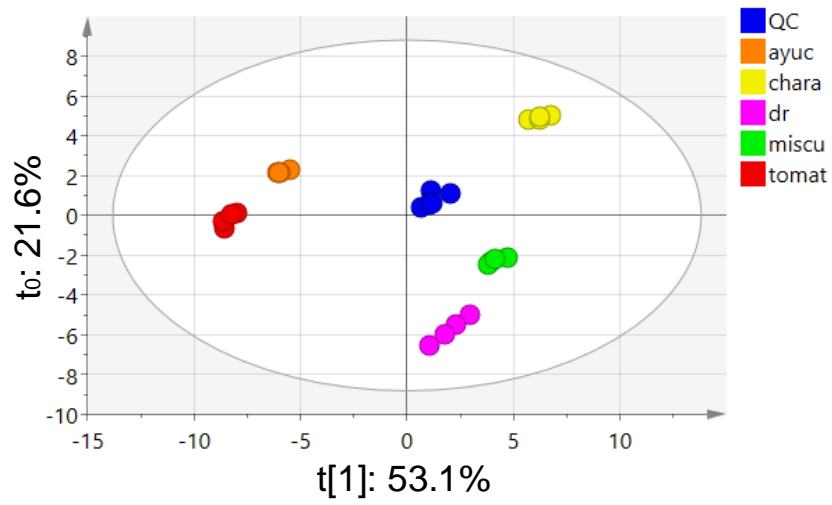


Figura 99. PCA(A) y OPLS-DA (B) de los perfiles UHPLC-HRMS de la localidad de Chinchá, Chiclayo y Huaral de la clase accesión. El PCA muestra un $R^2X=0.961$, el OPLS-DA muestra un $R^2Y=0.988$ y $Q^2=0.982$. CV-ANOVA ($p=0$). La varianza explicada es del 64.7 y 62.1%, respectivamente.

A)



B)

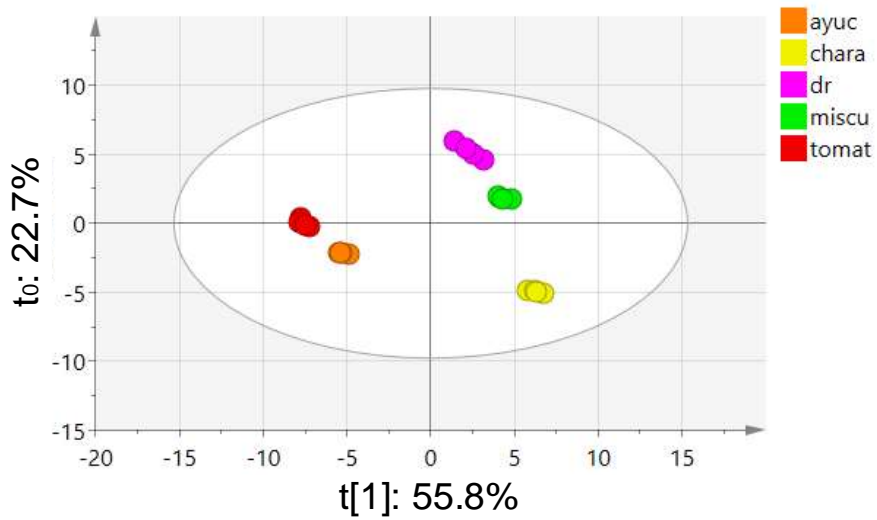


Figura 100. PCA (A) y OPLS-DA (B) de los perfiles UHPLC-HRMS de la localidad de Chincha de la clase accesión. El PCA muestra un $R^2X=0.948$, el OPLS-DA muestra un $R^2Y=0.995$ y $Q^2=0.991$. CV-ANOVA ($p=0$). La varianza explicada es del 74.7y 78.5%, respectivamente.

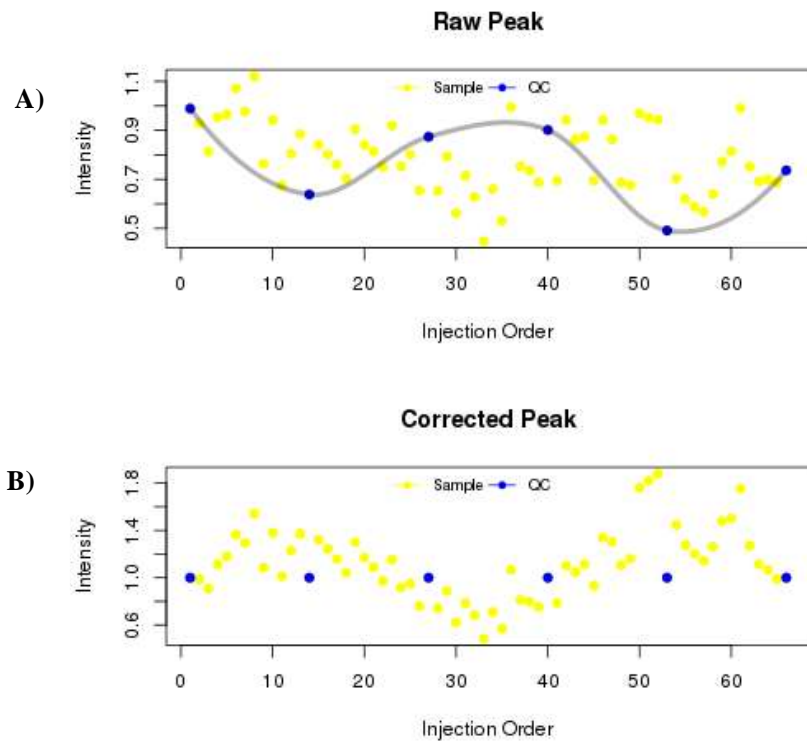
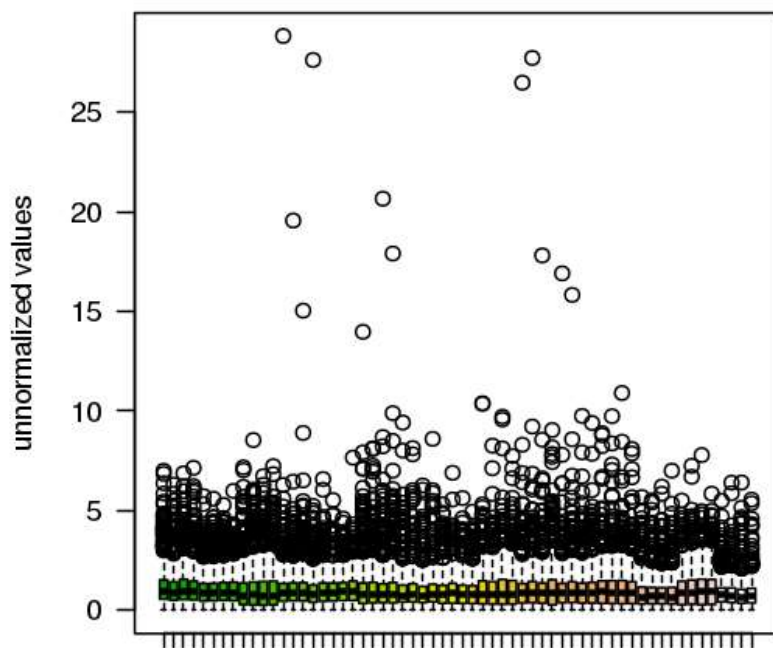


Figura 101. Corrección de picos cromatográficos basado en las muestras QC. En A se observa la data cromatográfica cruda inicial (6 QC). Luego, los picos son corregidos utilizando un ajuste de regresión polinomial. El ajuste permite obtener una intensidad de ajuste del QC constante mostrado en B.

A)



B)

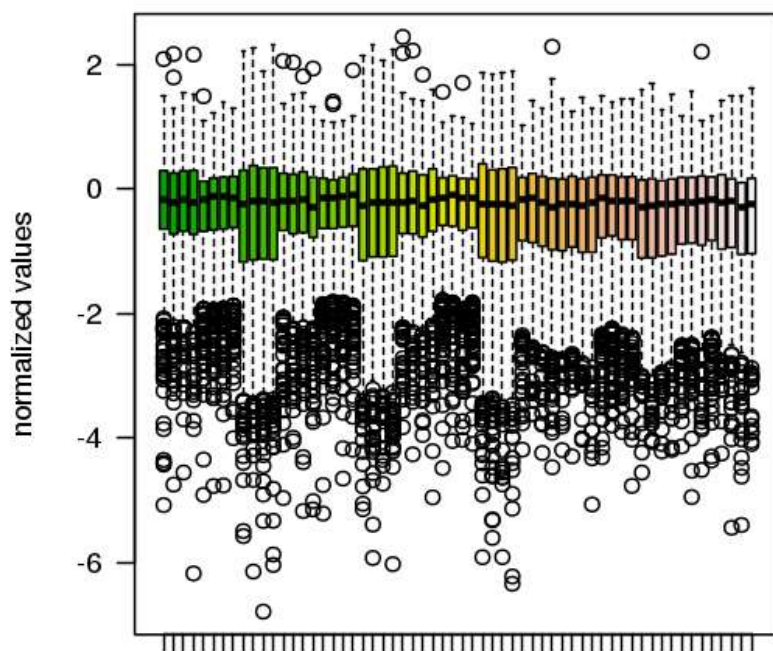


Figura 102. Normalización de la data cromatográfica utilizando el método de normalización EigenMS. En A se observa el diagrama boxplots de intensidades de la data cromatográfica antes de la normalización, en B se observa los diagramas boxplots de intensidades de la data normalizada, no se muestran las intensidades de los QC. Cada intensidad aparece en el orden de ejecución de la corrida.

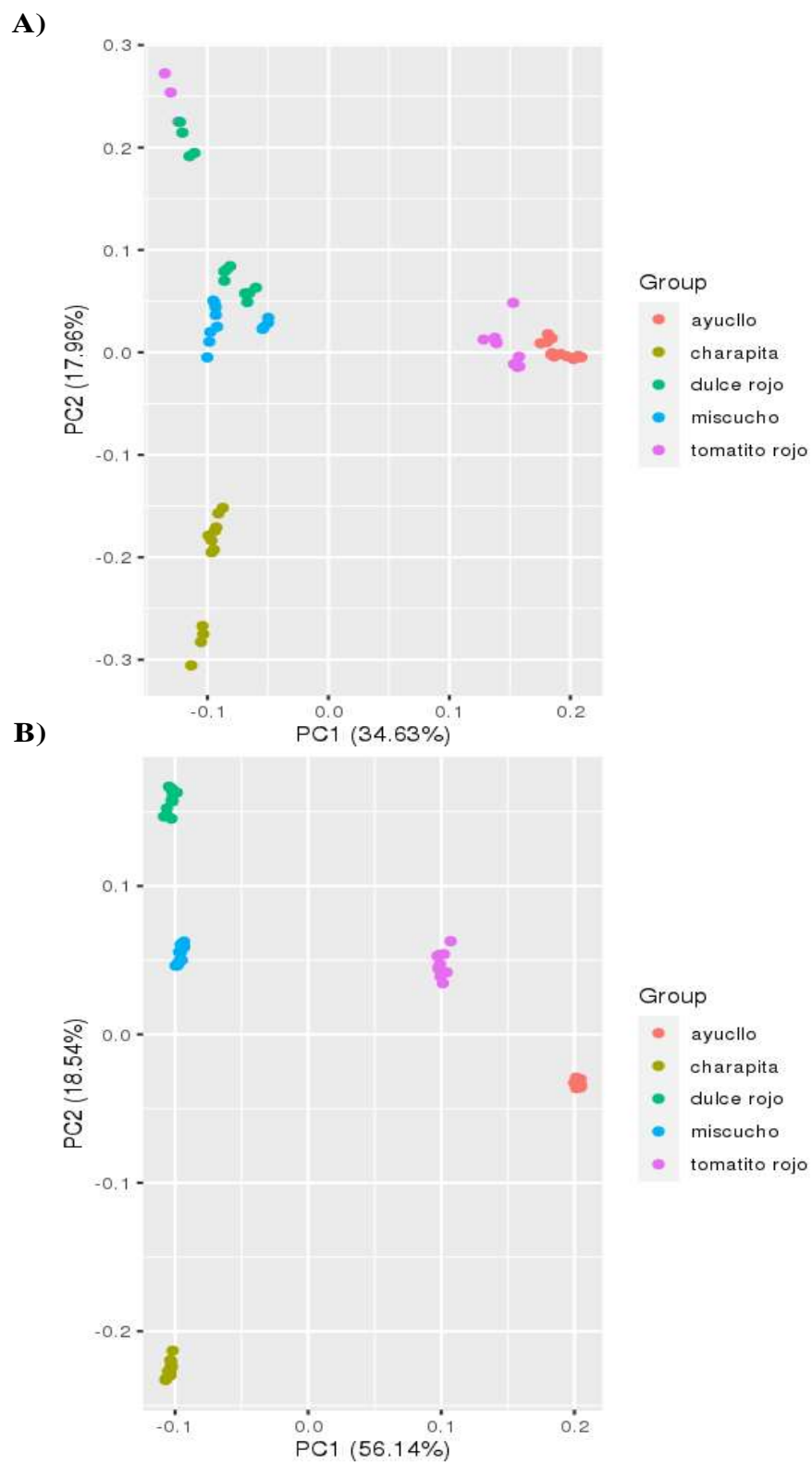


Figura 103. PCA de la clase accesión. Las figuras 103 A y 103 B muestran el PCA antes y después de la normalización por el método EigenMS.

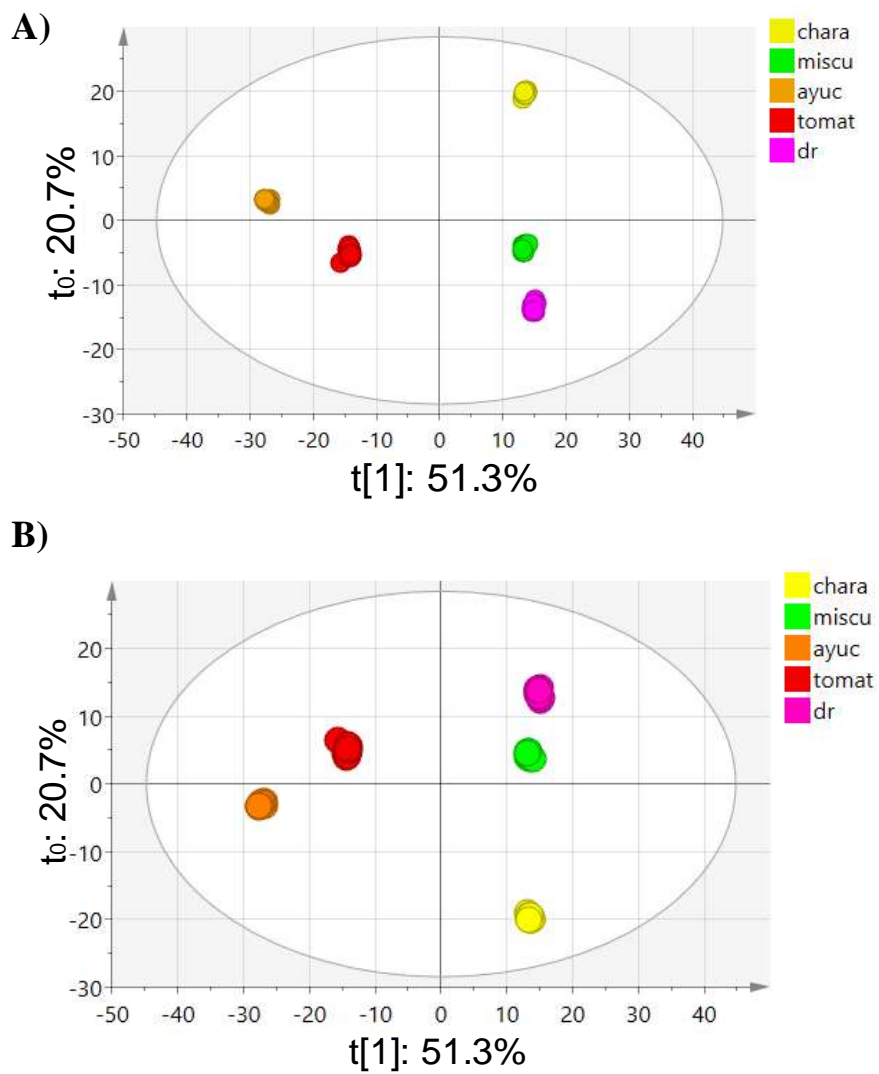


Figura 104. PCA (A) y OPLS-DA(B), data cromatográfica de los ajíes de las localidades de Chincha, Chiclayo y Huaral, normalizada por el método EigenMS de la clase accesión. El PCA muestra un $R^2_x = 0.909$, el OPLS-DA muestra un $R^2_Y = 0.996$; $Q^2 = 0.995$. CV-ANOVA ($p = 0$). La varianza explicada es del 72% para ambos modelos.

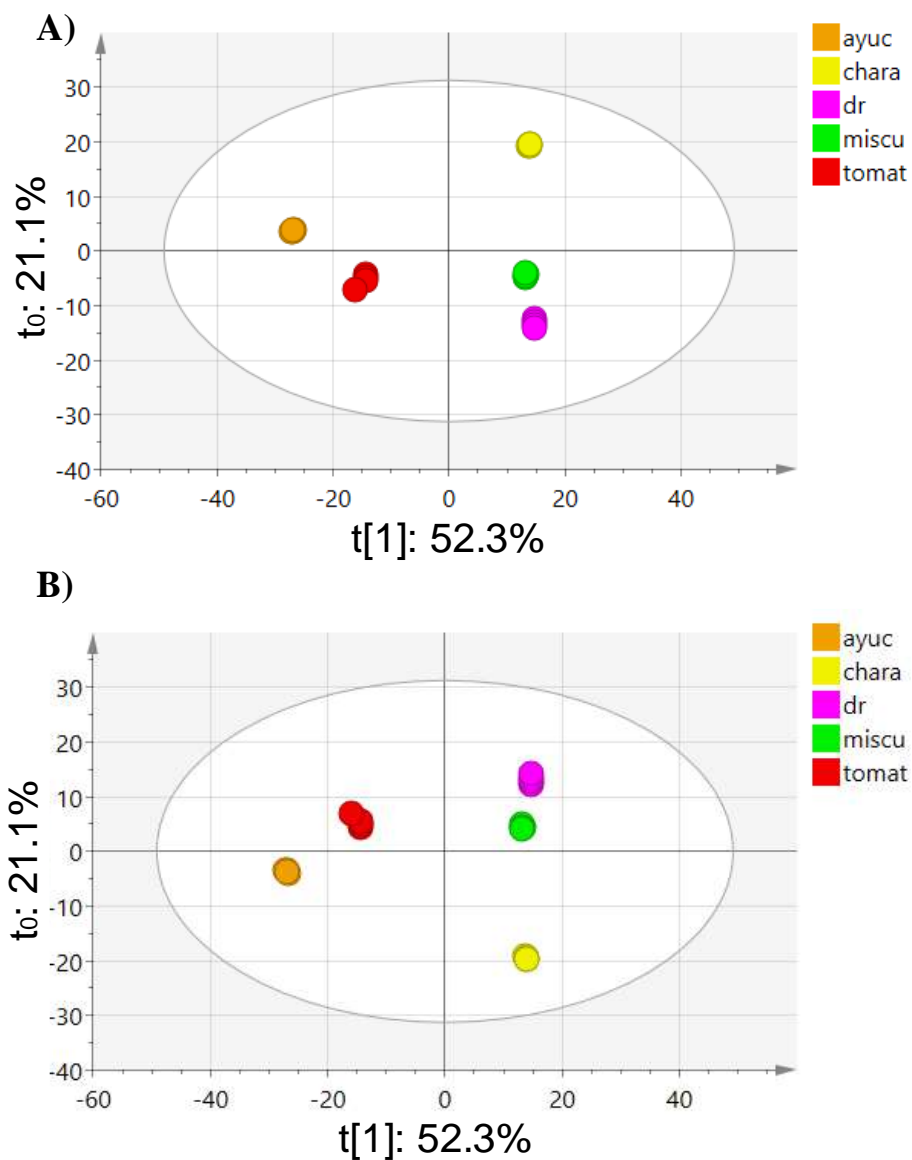


Figura 105. PCA (A) y OPLS-DA(B), data cromatográfica de los ajíes de la localidad Chíncha, normalizada por el método EigenMS de la clase accesión. El PCA muestra un $R^2X=0.904$, el OPLS-DA muestra un $R^2Y=0.997$; $Q^2=0.990$. CV-ANOVA ($p=0$). La varianza explicada es del 73.4% para ambos modelos.

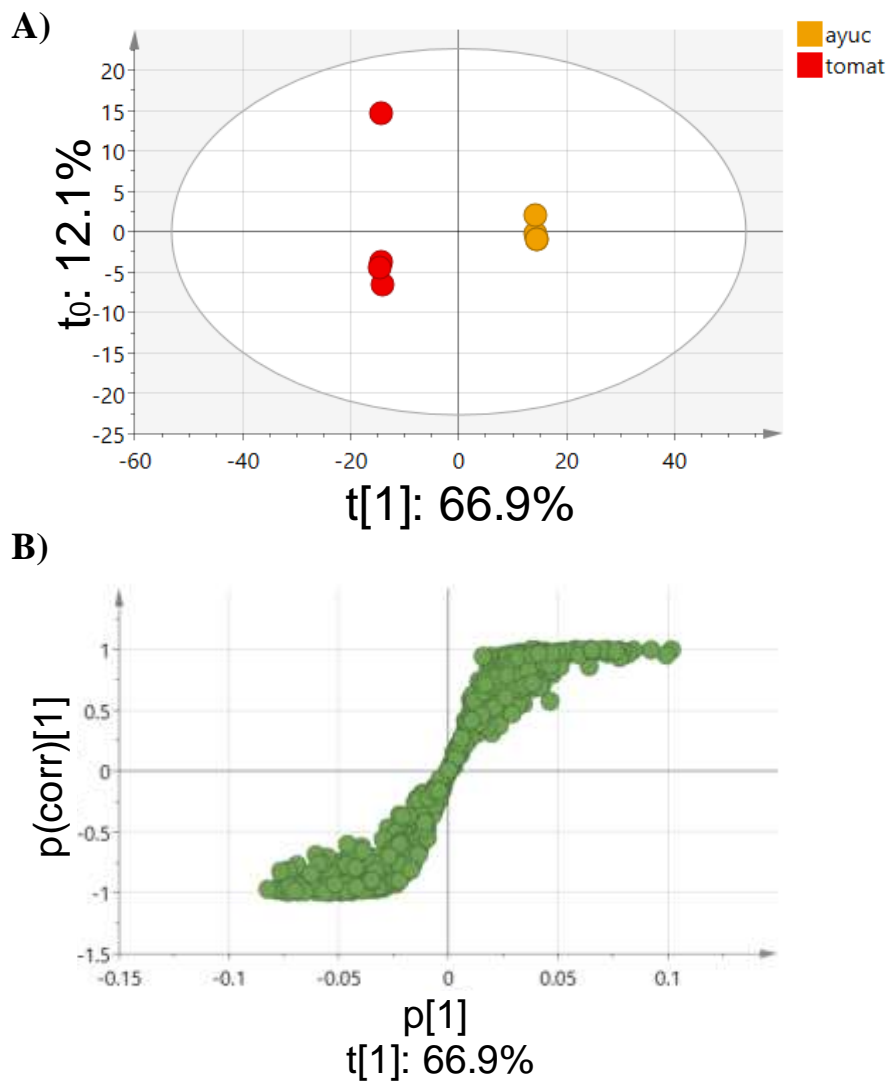


Figura 106. Diagramas OPLS-DA y S-plot. En A se observa el OPLS-DA ($R^2Y=1$ y $Q^2=0.994$) de los individuos *Tomatito* (*C.bac*) y *Ayucllo* (*C.bac*), en B se observa el diagrama de dispersión S-plot. Los puntos extremos del diagrama S-plot denotan las variables con la más alta covarianza (Biomarcadores).

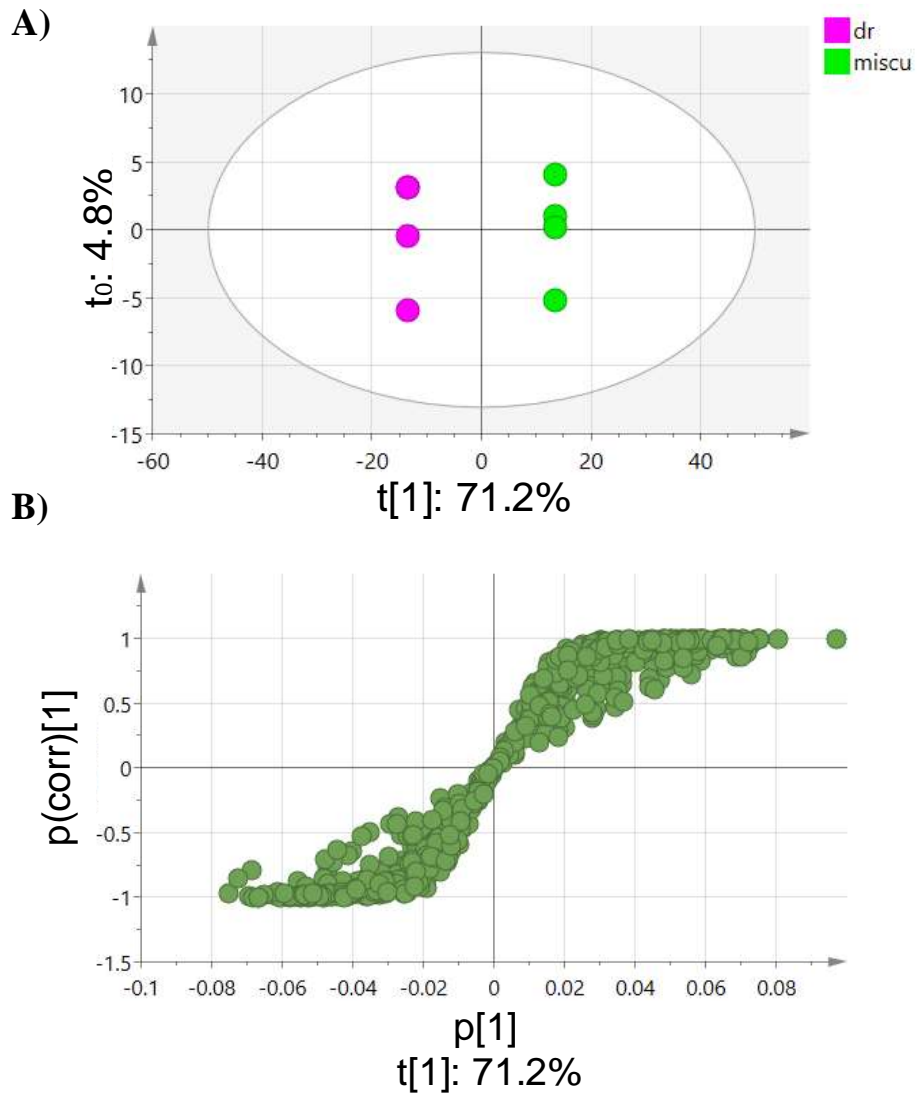


Figura 107. Diagramas OPLS-DA y S-plot. En A se observa el OPLS-DA ($R^2 = 1$ y $Q^2 = 0.994$) de los individuos *Dulce rojo* (*C.chin*) y *Miscucho* (*C.chin*). En B se observa el diagrama de dispersión S-plot. Los puntos extremos del diagrama S-plot denotan las variables con la más alta covarianza (Biomarcadores).

Tabla 39. Resultados del análisis S-plot, biomarcadores a partir del modelo OPLS-DA de las accesiones Dulce rojo (Dr) vs

Accesión	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(Score/Dot product/cosine)	p[1]	p(corr)[1]
Dr	POS	3.83	514.195	C26H30N2O9	5-Hydroxytryptamine; Nb-(4-Hydroxy-3-methoxy-E-cinnamoyl), O-β-D-glucopyranoside	Alcaloide	DNP-Solanaceae	7.0675	-0.0695	-0.9927
Dr	Neg	3.24	402.131	C21H22O8	3,4',5,7,8-Pentahydroxyflavanone; 8-Me ether, 3-pentanoyl	Flavonoide	DNP-Solanaceae	6.8241	-0.0681	-0.9993
Dr	Neg	3.96	466.126	C25H22O9	Silychristin; 3-Deoxy	Flavonoide	DNP-Solanaceae	7.3555	-0.0667	-0.9980
Dr	POS	4.42	372.214	C19H32O7	9-Hydroxy-5-megastigmen-4-one; 9-O-β-D-Glucopyranoside	Terpenoide	DNP-Solanaceae	7.3131	0.1734	0.9986
Dr	POS	0.64	308.159	C13H20N6O3	3'-Amino-3'-deoxyadenosine; 3',6,6-Tri-N-Me	Nucleósido	DNP-Solanaceae	7.2594	0.1196	0.8805
Dr	POS	0.68	295.128	C11H21NO8	3,4-Dihydroxy-2-(hydroxymethyl)pyrrolidine; 4-O-β-D-Glucopyranoside	Alcaloide	DNP-Capsicum	6.5659	0.1265	0.9333
Dr	POS	1	293.147	C12H23NO7	N-Fructosyl isoleucine	Aminoácido	MS-DIAL	1000	0.1376	0.9936
Dr	Neg	4.42	418.199	C23H30O7	3,10-Dihydroxy-4,15,17-sphenobatrien-12,13-olid-19-oic acid; 10-Ac, Me ester	Terpenoide	DNP-Solanaceae	6.468	0.1595	0.9991
Miscu	Neg	4.55	1118.499	C49H82O28	2,6,10-Farnesatriene-1,12-diol; Bis-O-[α-L-rhamnopyranosyl-(1→2)-[β-D-xylopyranosyl-(1→3)]-β-D-glucopyranoside]	Terpenoide	DNP-Solanaceae	6.868	-0.1254	-0.9961
Miscu	Neg	0.68	372.121	C20H20O7	Sigmoidin C; 8'-Deoxy, 3',4'-dihydro, 3'ξ,4'ξ-dihydroxy	Flavonoide	DNP-Solanaceae	7.3904	-0.0464	-0.9662
Miscu	Neg	4.06	510.174	C24H30O12	1,5-Bis(3,4-dihydroxyphenyl)-4,5-dihydroxy-1-pentanone; 7'-Me ether, 8'-O-β-D-glucopyranoside	Aromático	DNP-Solanaceae	7.0658	-0.0690	-0.9564
Miscu	Neg	4.07	442.1839	C21H30O10	3-Phenyl-2-propen-1-ol; O-[α-L-Rhamnopyranosyl-(1→6)-β-D-glucopyranoside]	Aromático	DNP-Capsicum	7.5662	-0.0668	-0.9964
Miscu	POS	2.04	250.132	C13H18N2O3	N-Cinnamoylputrescine; 3'',4''-Dihydroxy	Alcaloide	DNP-Solanaceae	7.7706	-0.0651	-0.9976
Miscu	POS	2.09	250.121	C14H18O4	2,4-Pentanediol; 4-O-(4-Hydroxy-E-cinnamoyl)	Alifático	DNP-Solanaceae	7.6988	-0.0648	-0.9192
Miscu	Neg	3.71	440.168	C21H28O10	1,2-Cyclohexanediol; O-[3,4-Dihydroxy-E-cinnamoyl-(→2)-β-D-glucopyranoside]	Alifático	DNP-Capsicum	7.335	-0.0550	-0.9864
Miscu	POS	2.9	264.147	C14H20N2O3	N-Cinnamoylputrescine; 3''-Methoxy, 4''-hydroxy	Alcaloide	DNP-Capsicum	7.2171	-0.0569	-0.9949

Miscucho

Tabla 40. Resultados del análisis S-plot, biomarcadores a partir del modelo OPLS-DA de las accesiones Tomatito rojo vs Ayuello

Accesión	ESI	RT	Masa exacta	Fórmula molecular	Metabolito	Clase química	Encontrado en :	(score/Dot product/cosine)	p[1]	p(corr)[1]
Tomatito	POS	5.71	634.226	C31H38O14	3-(4-Hydroxyphenyl)-2-propen-1-ol; 1-O-[α -L-Rhamnopyranosyl-(1 \rightarrow 3)-[4-hydroxy-3-methoxy-E-cinnamoyl-(\rightarrow 4)]- β -D-glucopyranoside]	Aromático	DNP-Capsicum	6.4032	-0.0705	-0.9700
Tomatito	Neg	3.94	598.168	C30H30O13	4-Hydroxybenzyl alcohol; α -O-(3,4-Dihydroxybenzoyl), 4-O-[4-hydroxy-3-methoxy-E-cinnamoyl-(\rightarrow 6)]- β -D-glucopyranoside]	Aromático	DNP-Capsicum	8.0208	-0.0486	-0.9762
Tomatito	Neg	3.84	434.100	C24H18O8	3,5,7-Trihydroxyflavanone; 3-O-(3,4-Dihydroxy-E-cinnamoyl)	Flavonoide	DNP-Solanaceae	7.4717	-0.0526	-0.9662
Tomatito	Neg	3.54	342.095	C15H18O9	3-(3,4-Dihydroxyphenyl)-2-propenoic acid; 4'-O-?-D-Glucopyranoside	Aromático	DNP-Capsicum	7.3904	-0.0442	-0.9622
Tomatito	Neg	8.49	596.283	C30H44O12	7-Megastigmene-3,6,9-triol; 3-O-[4-Hydroxy-3,5-dimethoxy-E-cinnamoyl-(\rightarrow 6)]- β -D-glucopyranoside]	Terpenoide	DNP-Capsicum	7.0886	-0.0596	-0.9138
Tomatito	Neg	9.12	562.314	C31H46O9	2,3,14,20,22,28-Hexahydroxy-6-oxostigmast-7-en-26-oic acid; 26?28-Lactone, 22-Ac	Esteroides	DNP-Solanaceae	7.3568	-0.0317	-0.9590
Tomatito	POS	8.43	700.367	C35H56O14	1,2,3,14,20,22,25-Heptahydroxycholest-7-en-6-one; 25-Ac, 22-O- β -D-galactopyranoside	Esteroides	DNP-Solanaceae	7.4604	-0.0574	-0.9017
Tomatito	Neg	8.43	724.367	C37H56O14	3,14,16-Trihydroxycard-20(22)-enolide; 16-Ac, 3-O-[?-D-glucopyranosyl-(1 \rightarrow 4)-2,6-dideoxy- β -D-ribohexopyranoside]	Esteroides	DNP-Solanaceae	7.1729	-0.0570	-0.8269
Ayuello	POS	10.24	255.256	C16H33NO	2-Amino-9-hexadecen-3-ol; (2S,3R,9Z)-form	Alcaloide	DNP-Capsicum	7.3465	0.0841	0.9981
Ayuello	Neg	3.62	472.143	C17H28O15	a-D-Glucopyranuronosyl-(1 \rightarrow 2)- β -D-xylopyranosyl-(1?4)-D-xylose; O4"-Me	Carbohidrato	DNP-Solanaceae	8.0123	0.0819	0.9595
Ayuello	POS	4.27	594.158	C27H30O15	7-[(2S-,3R,4S,5S,6R)-6-[[[(2S,3R,4R)-3-4-dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxymethyl]-3-4-5-trihydroxyoxan-2-yl]oxy-5-hydroxy-3-(4-hydroxyphenyl)-6-methoxychromen-4-one	Flavonoide	GNPS	0.76	0.0738	0.9733
Ayuello	Neg	3.5	518.142	C25H26O12	Edgeworic acid; Me ester, 7-O-?-D-glucopyranoside	Cumarina	DNP-Solanaceae	7.567	0.0721	0.9664
Ayuello	Neg	4.28	640.143	C31H28O15	Quercetin 3-glycosides; 3-O-[4-Hydroxy-3-methoxy-E-cinnamoyl-(\rightarrow 6)]- β -D-galactopyranoside]	Flavonoide	DNP-Solanaceae	7.4872	0.0717	0.9920
Ayuello	POS	10.44	281.272	C18H35NO	2-Octadecenoic acid; Amide	Alcaloide	DNP-Solanaceae	7.7889	0.0689	0.9844
Ayuello	POS	11.03	283.288	C18H37NO	16-Methylheptadecanoic acid; Amide	Alcaloide	DNP-Capsicum	7.4352	0.0654	0.9962
Ayuello	POS	7.18	293.199	C17H27NO3	Capsaicin; 7-Methyloctanoyl analogue	Alcaloide	DNP-Capsicum	6.1108	0.0548	0.9599

En la figura 108 se observa el perfil metabolómico en el diagrama Heatmap de la clase accesión de los ajíes de la localidad de Chincha construido a partir del modelo PLS-DA de las 824 características VIP (variables de importancia). La imagen coloreada comunica claras diferencias entre las 5 accesiones de ajíes Ayuclo (ayuc), Charapita (chara), Dulce rojo (dr), Miscucho (miscu), y Tomatito rojo (tomat), a través del agrupamiento jerárquico (HCA) realizado tanto para filas y columnas en la matriz de datos. En las filas se observa el agrupamiento jerárquico por accesiones y en columnas el agrupamiento jerárquico de los metabolitos. Las observaciones han sido ordenadas de acuerdo a su similaridad. La escala de colores muestra los valores altos y bajos con valores similares uno cerca del otro. Como distancia de medida se utilizó la distancia Euclidiana y como algoritmo de agrupamiento se utilizó el algoritmo Ward.

En las figuras 109-112 se observan los diagramas Boxplots de los principales metabolitos biomarcadores identificados para las accesiones Ayuclo (*C.bac*), Tomatito rojo (*C.bac*), Dulce rojo (*C. chin*) y Miscucho (*C. chin*) mediante análisis S-plot, específicos para cada accesión. Dichos metabolitos son específicos puesto que están presentes en una accesión dada en altas concentraciones mientras que en otras accesiones es inexistente o está en muy baja concentración.

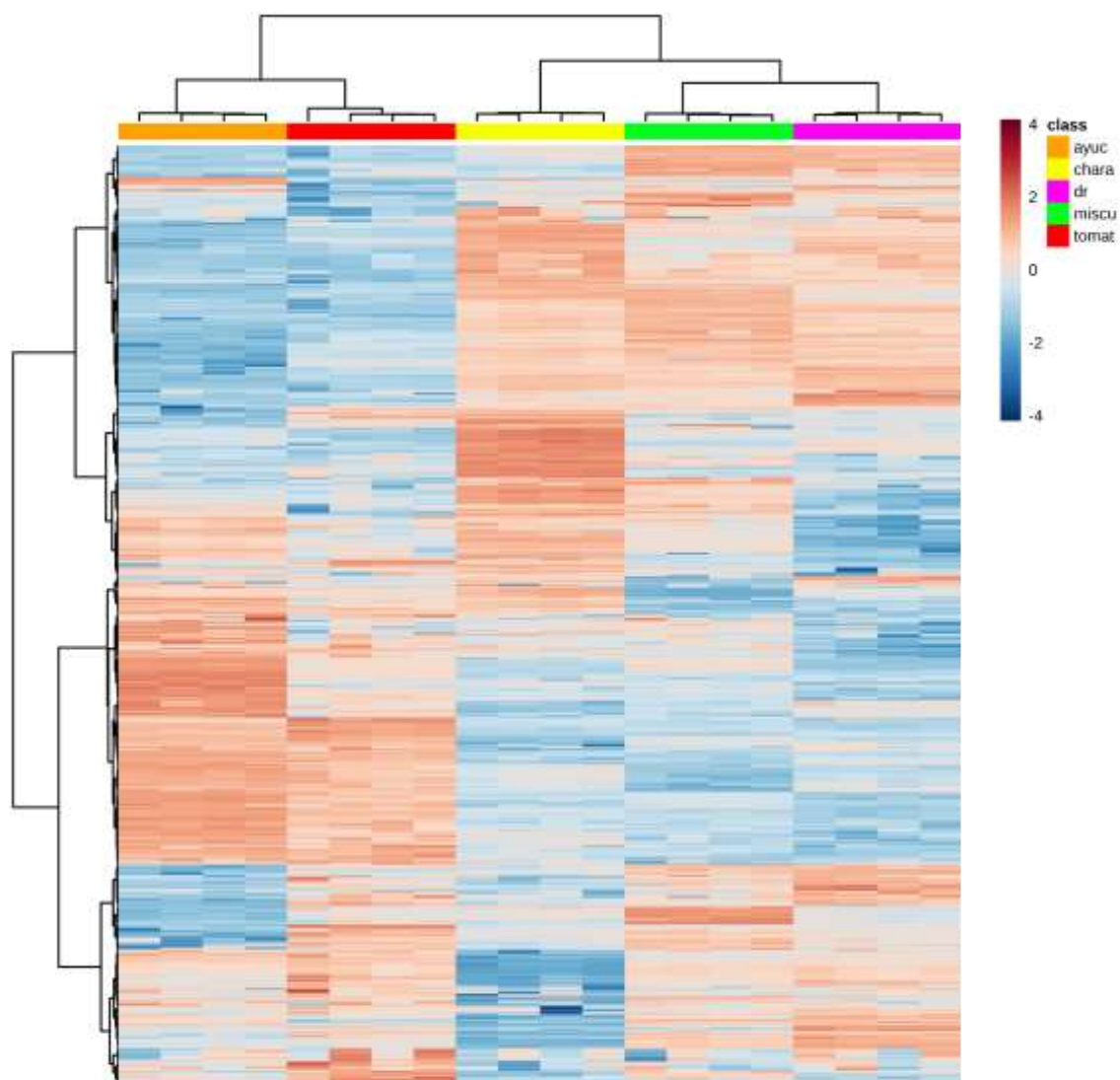


Figura 108. Heatmap de las 824 características de metabolitos no volátiles a partir del modelo PLS-DA de la clase accesión del fruto entero de ají de la localidad de Chincha.

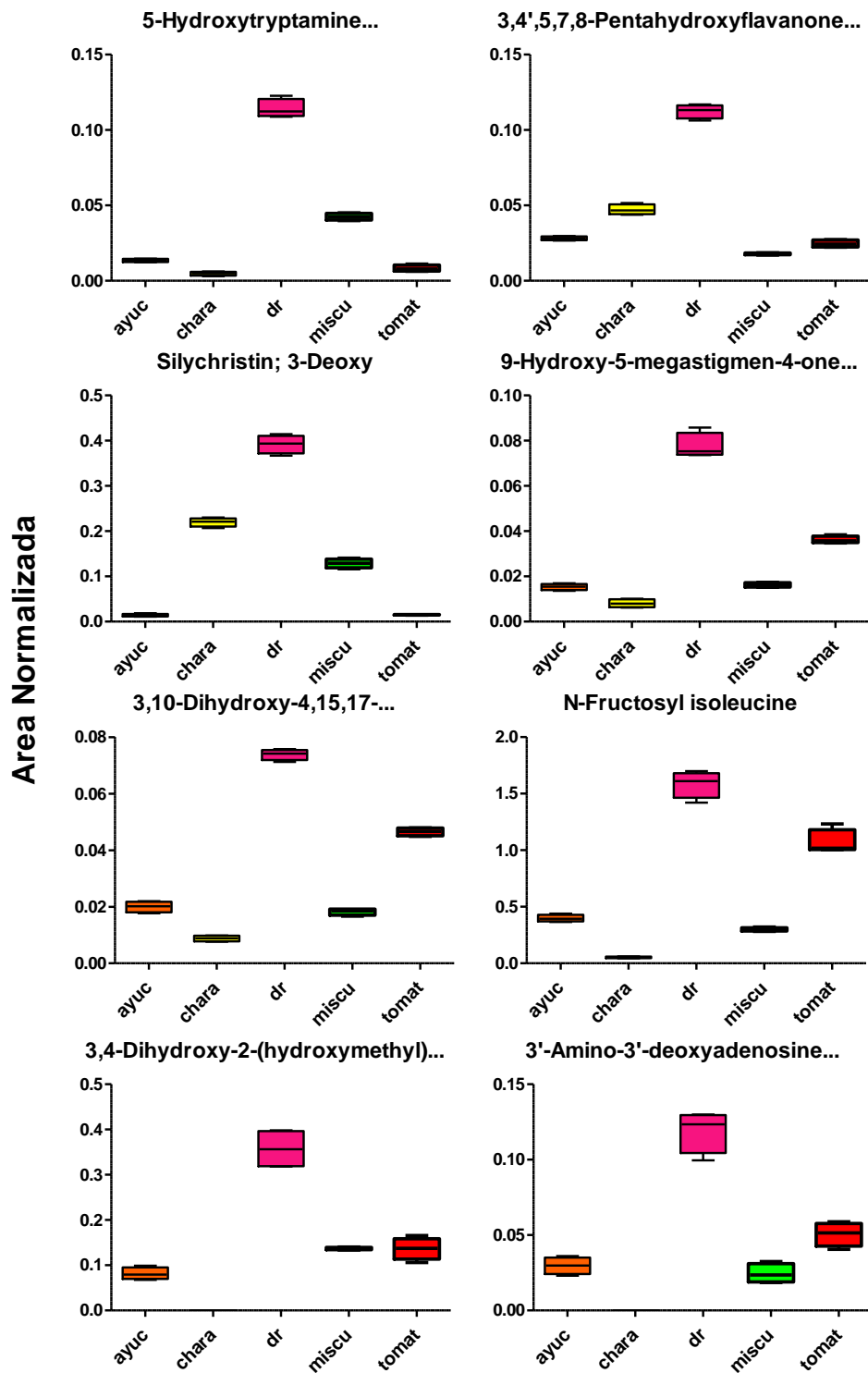


Figura 109. Boxplots de los metabolitos biomarcadores de la accesión Dulce rojo versus Miscucho

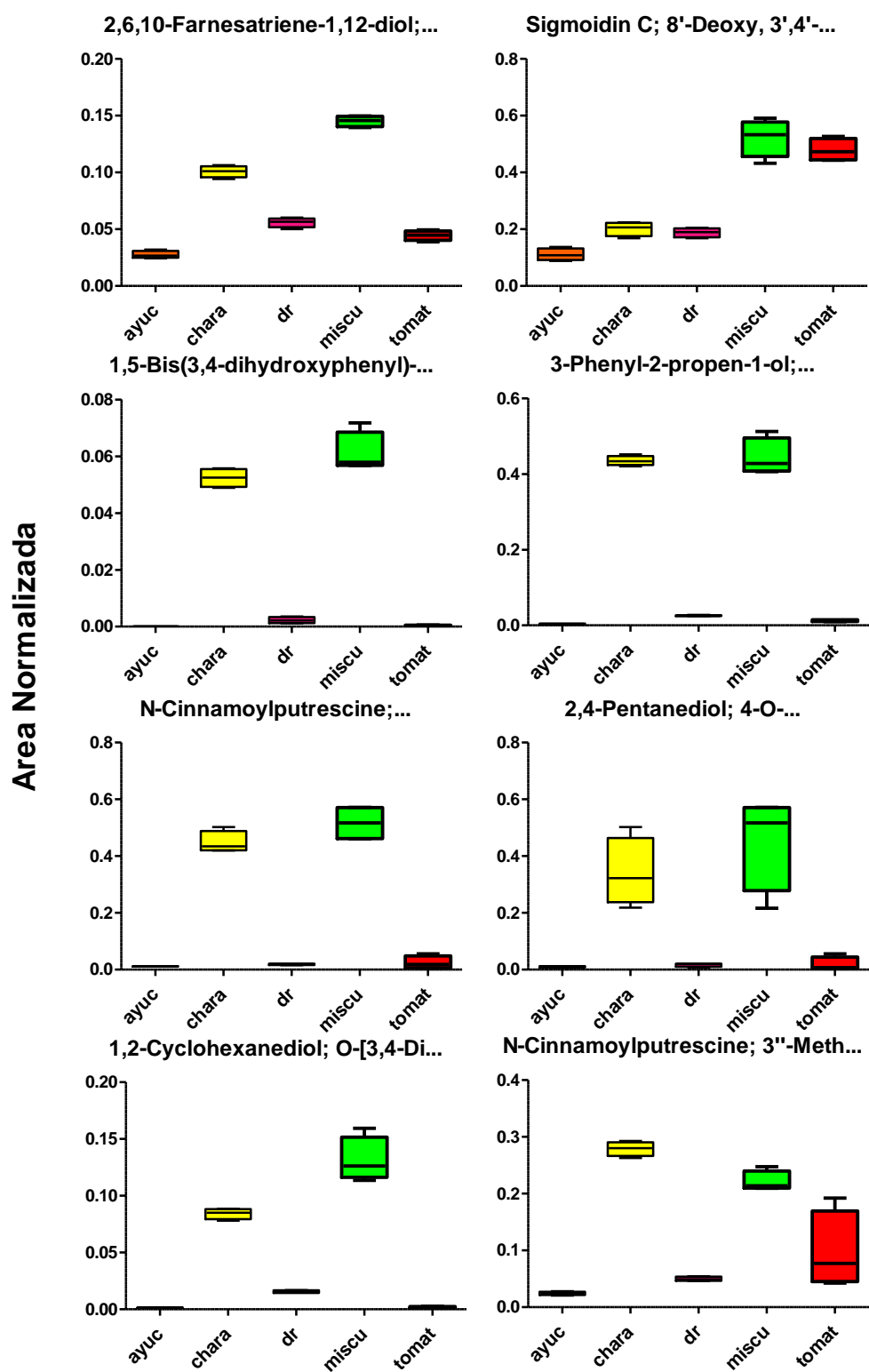


Figura 110. Boxplots de los metabolitos biomarcadores de la accesión Miscucho versus Dulce rojo

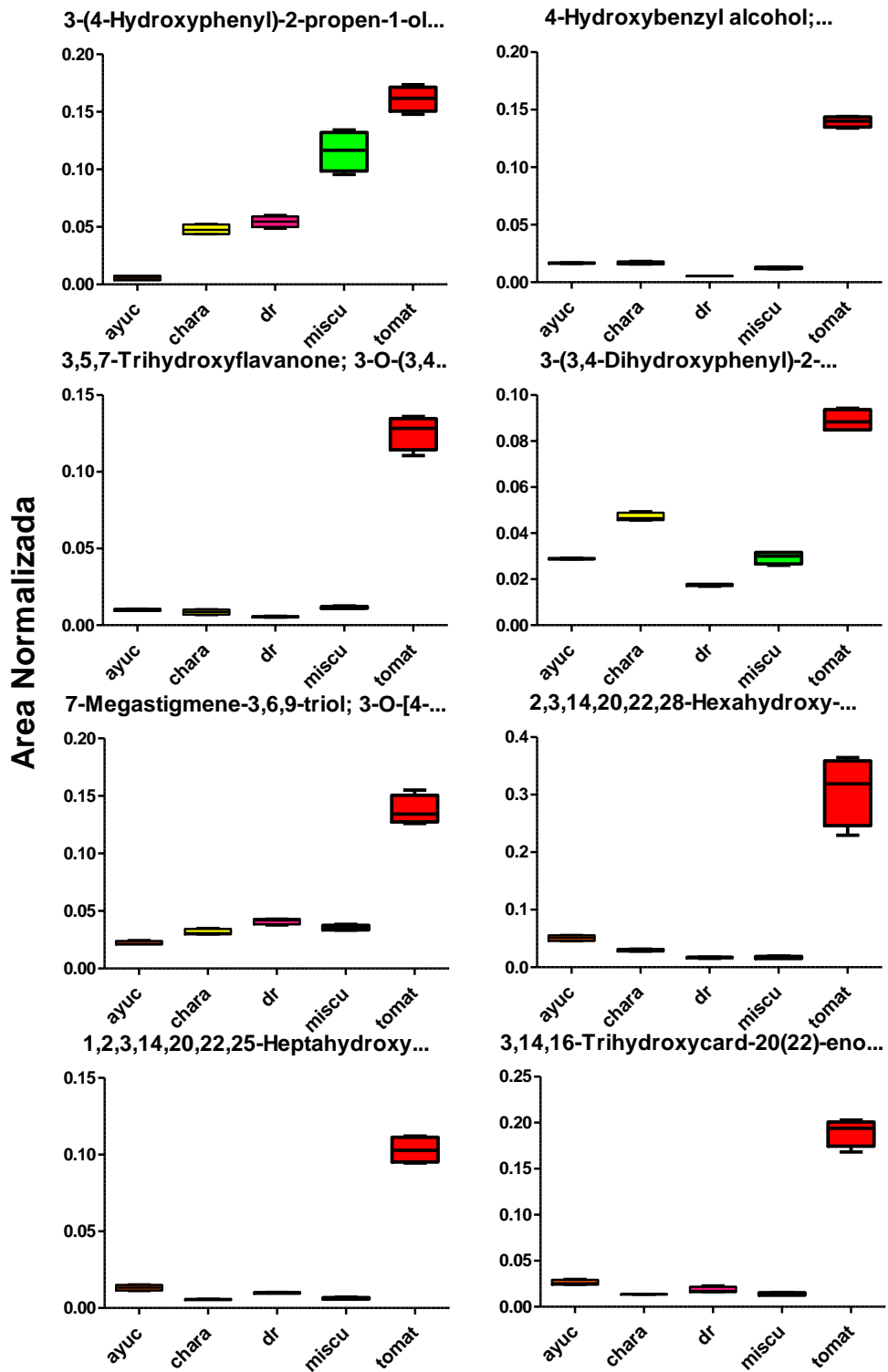


Figura 111. Boxplots de los metabolitos biomarcadores de la accesión Tomatito rojo versus Ayucllo

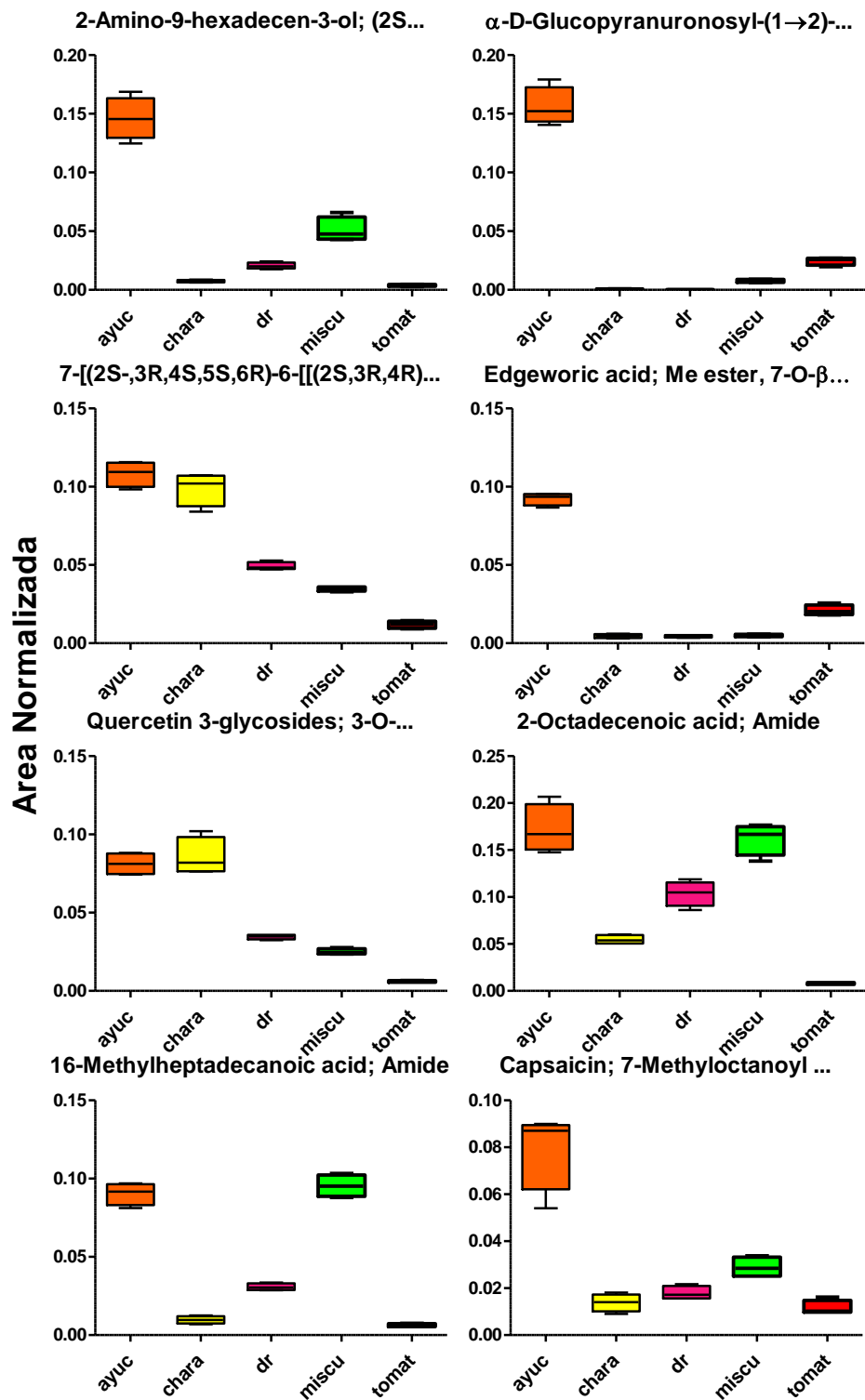


Figura 112. Boxplots de los metabolitos biomarcadores de la accesión Ayucllo versus Tomatito rojo

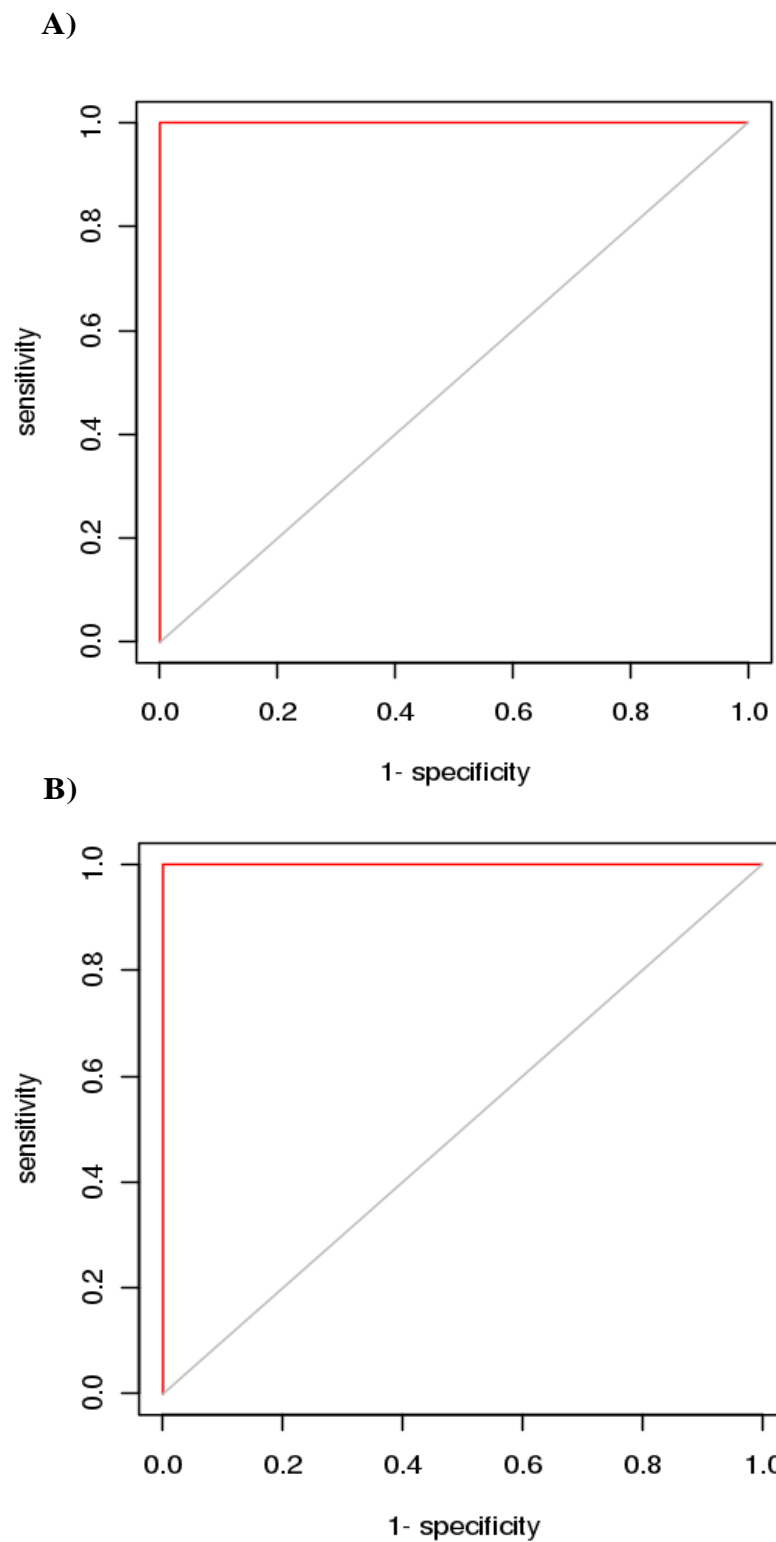


Figura 113. Curva ROC para la evaluación de la influencia del método de normalización EigenMS en la precisión de clasificación de la clase especies (**A**) y accesiones (**B**), el valor de AUC = 1.

7.7. Redes moleculares y clases químicas de metabolitos

Las figuras 114 y 115 muestran el análisis de red molecular basado en la similaridad de los patrones de fragmentación MS2. Estas redes fueron diseñadas a partir de la data MS2 proporcionada por el software MS-DIAL (archivo *.mgf (GNPS)). Posteriormente se utilizó el software MetGem para diseñar la red molecular y hacer anotaciones adicionales utilizando la base de datos GNPS. El software Cytoscape permitió importar la red molecular generada por MetGem e introducir información adicional tal como la clase química de los metabolitos, los valores VIP del modelo OPLS-DA dados por SIMCA P y promedio de áreas relativas para desarrollar gráficos Pie. Las moléculas con patrones similares de MS2 se agrupan formando cluster. Las líneas de unión (Edges) denotan la fuerza de la similaridad (cosine), mientras que las esferas denotan los metabolitos (Nodos). Los Nodos que se unen a Edges con líneas más gruesas denotan alta similaridad. La Red Molecular es una herramienta importante que permite hacer anotaciones adicionales de metabolitos no identificados por similaridad MS2 con otros patrones MS2 de moléculas ya identificadas, y tentativamente elucidar nuevos biomarcadores.

Al analizar los metabolitos de acuerdo a sus clases químicas, en el modo de ionización ESI (+) se encontraron 22 clases químicas. Entre las más importantes, destacan un total de 71 Alcaloides, 33 terpenoides, 32 flavonoides y 20 carbohidratos como los más abundantes, con 185 metabolitos no identificados (tabla 38). Mientras que en el modo de ionización ESI (-) se encontraron 18 clases químicas. Entre las más importantes, se observó que existen un total de 6 alcaloides,

56 terpenoides, 55 flavonoides y 32 carbohidratos, con 263 metabolitos no identificados (tabla 39). Entre los metabolitos menos abundantes en ESI (+) se encontraron 8 Lignanos, 1 nucleósido y 1 cumarina. Mientras que en ESI (-) se encontraron 12 lignanos, 1 nucleósido y 3 cumarinas, entre otros.

De la especie *C. frutescens* se identificaron 8 biomarcadores de los cuales existen dos flavonoides, dos compuestos aromáticos, un esteroide, un alcaloide, un nucleósido y un terpenoide. Se ha observado que del total de metabolitos en el modo de ionización ESI (-) solo existe un nucleósido. Este nucleósido identificado como **5-tioadenosina** es uno de los fuertes candidatos de biomarcador para la especie *C. frutescens*. De la especie *C. chinense* se identificaron cuatro terpenoides, un carbohidrato, dos esteroides y un lignano. El lignano identificado como **3,3',4,4'-Tetrahydroxy-9-nor-8,8'-lign-7'-en-9',7-olide; 3'-Me** encontrado en el modo de ionización ESI (-) es uno de los fuertes candidatos de biomarcadores para la especie *C. chinense*. De la especie *C. baccatum*, existen cinco flavonoides, un compuesto aromático, una cumarina y un carbohidrato. La cumarina identificada como **Arteminorin B** es un metabolito encontrado en el modo de ionización ESI (-) y es uno de los fuertes candidatos de biomarcadores para esta especie.

Para la discriminación por accesiones se confrontaron las accesiones Dulce rojo con Miscucho, ambos de la especie *C. chinense* y Ayucillo con Tomatito rojo, ambos de la especie *C. baccatum*. De la accesión Dulce rojo se identificaron dos alcaloides, dos flavonoides, dos terpenoides, un nucleósido y un aminoácido como

biomarcadores. El nucleósido identificado como **3'-Amino-3'-deoxyadenosine**; **3',6,6-Tri-N-Me** y el aminoácido **N-Fructosyl isoleucine** son dos de los fuertes metabolitos candidatos de biomarcadores para la accesión Dulce rojo. De la accesión Miscucho se identificaron un terpenoide, un flavonoide, dos compuestos aromáticos, dos alcaloides, y dos compuestos alifáticos. Los alcaloides **N-Cinnamoylputrescine**; **3''-Methoxy, 4''-hydroxy** y **N-Cinnamoylputrescine**; **3'',4''-Dihydroxy** son dos de los fuertes candidatos de biomarcadores para la accesión Miscucho.

De igual manera, en la accesión Tomatito rojo se han identificado tres compuestos aromáticos, un flavonoide, un terpenoide y tres esteroides. Para esta accesión el flavonoide identificado como 3,5,7-Trihydroxyflavanone; 3-O-(3,4-Dihydroxy-E-cinnamoyl) encontrado en el modo de ionización ESI (-) es uno de los fuertes candidatos de biomarcadores. Mientras que para la accesión ayucillo, se identificaron cuatro alcaloides, dos flavonoides, un carbohidrato y una cumarina. Por su clase química, la cumarina **Edgeworic acid; Me ester, 7-O-β-D-glucopyranoside** es uno de los fuertes candidatos de biomarcadores para esta accesión.

En las figuras 116 y 117 los nodos muestran las áreas relativas en gráficos Pie. Estos gráficos permitieron identificar 4 metabolitos biomarcadores de las tres especies.

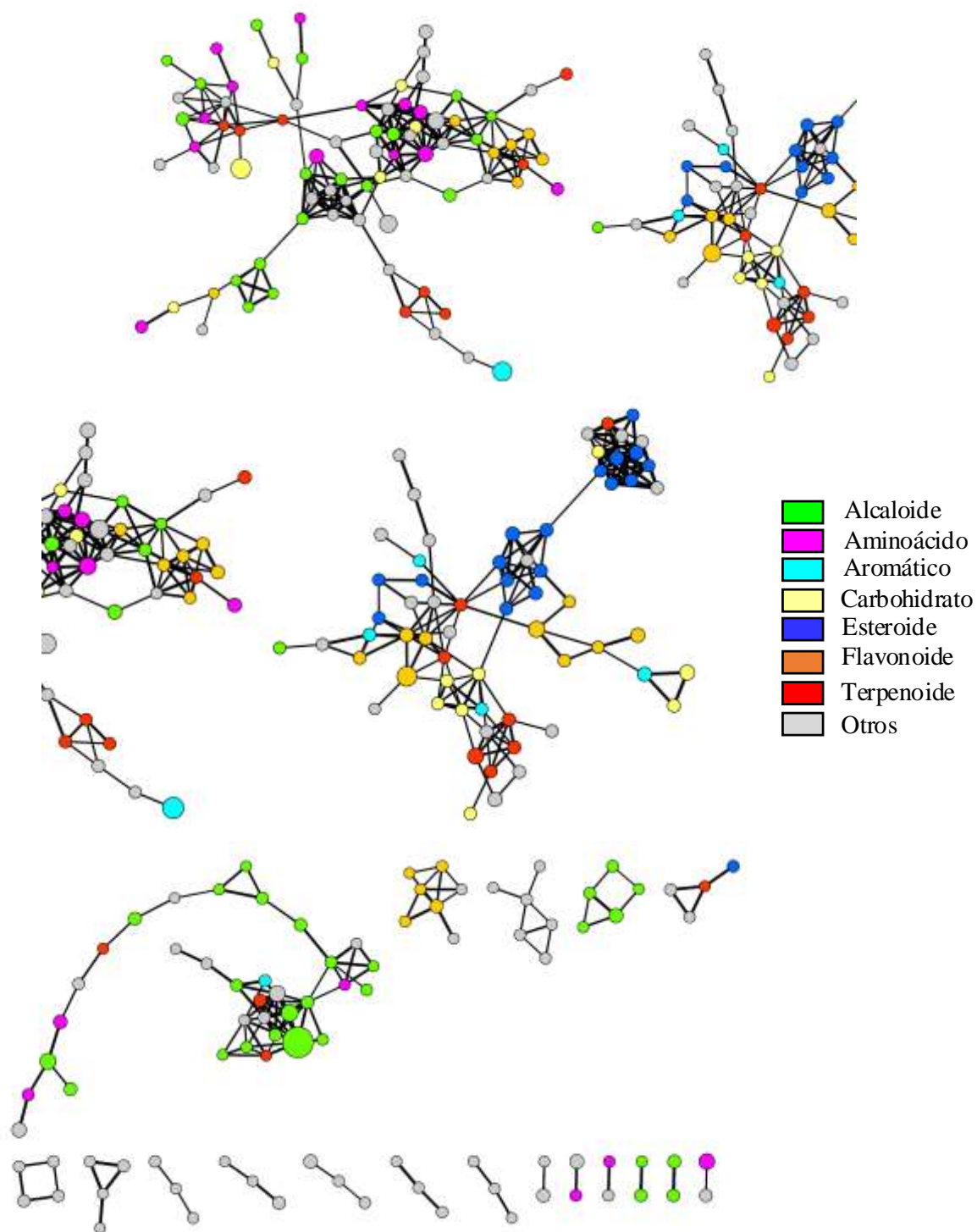


Figura 114. Red molecular de similitud ESI (+), (Molecular Networking) a partir de los patrones de fragmentación MS2. La red molecular fue construida utilizando el software MetGem y Cytoscape. Los colores de los nodos (esferas en colores) indican el tipo de clase química del metabolito presente. El volumen del nodo indica el valor de importancia (VIP). Altos volúmenes indican altos valores de VIP.

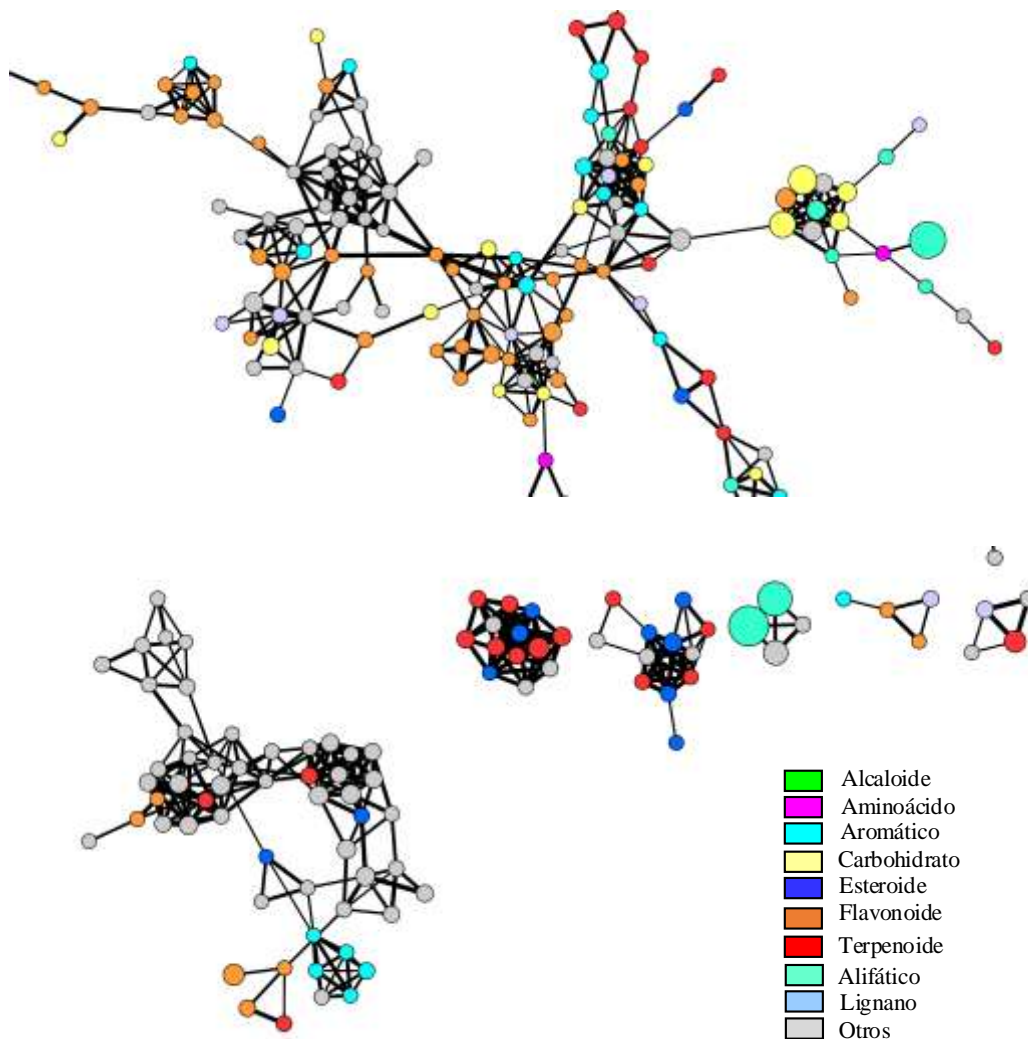


Figura 115. Red molecular de similaridad ESI (-), (Molecular Networking) a partir de los patrones de fragmentación MS2. La red molecular fue construida utilizando el software MetGem y Cytoscape. Los colores de los nodos (esferas en colores) indican el tipo de clase química del metabolito presente. El volumen del nodo indica el valor de importancia (VIP). Altos volúmenes indican altos valores de VIP.

Tabla 41. Número de metabolitos según sus clases química para el modo de ionización ESI (+)

Clase química	Número de metabolitos	% Clase química
Alcaloides	71	15.6
Alifáticos	21	4.6
Aminoácidos	26	5.7
Aromáticos	10	2.2
Benzofurano	2	0.4
Benzopirano	2	0.4
Benzoquinona	3	0.7
Benzotiazol	1	0.2
Carbohidrato	20	4.4
Cromona	1	0.2
Cumarina	1	0.2
Ditolano	1	0.2
Esteroides	22	4.8
Flavonoides	32	7.0
Glicerofosfocolina	3	0.7
Glicerofosfoetanolamina	2	0.4
Lignanos	8	1.8
Nucleósido	3	0.7
Polikétido	5	1.1
Terpenoide	34	7.5
Tioester	1	0.2
Tiomorfolina	1	0.2
No identificados	185	40.7
Total de Metabolitos	454	100

Diagrama circular de metabolitos no volátiles

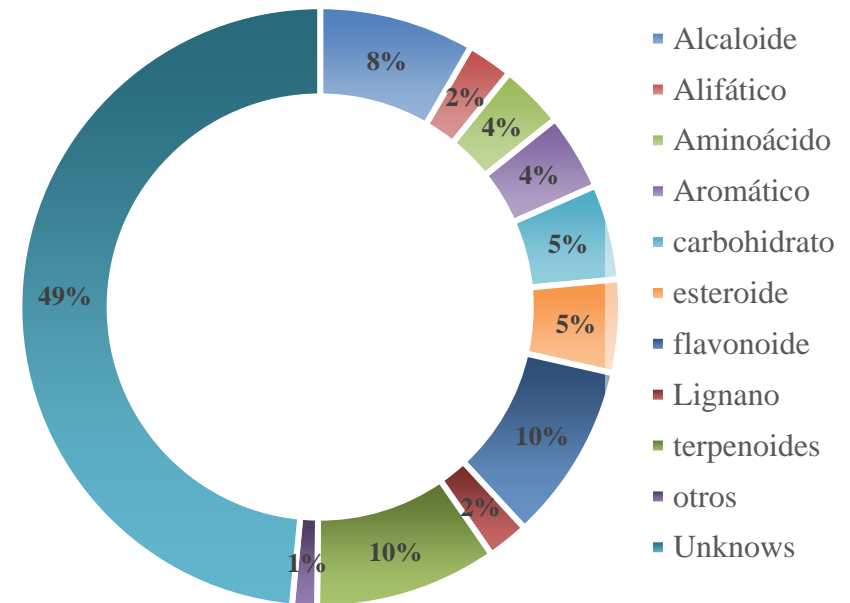
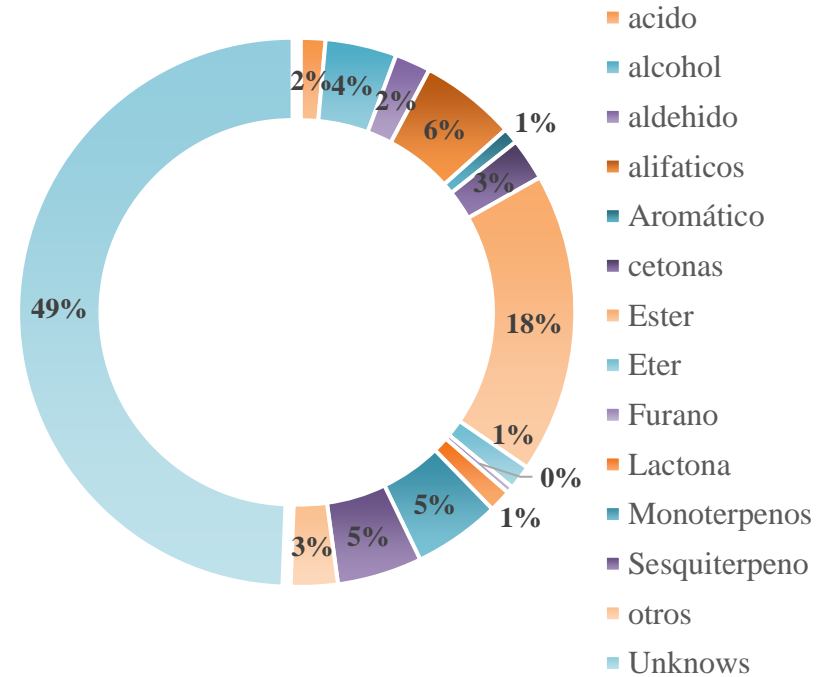


Tabla 42. Número de metabolitos según sus clases química para el modo de ionización ESI (-)

Clase química	Número de metabolitos	% Clase química
Alcaloides	6	1.2
Alifáticos	8	1.6
Aminoácidos	6	1.2
Aromáticos	28	5.6
Benzofurano	1	0.2
Benzoquinona	0	0.0
Benzotiazol	0	0.0
Benzopirano	2	0.4
Carbohidrato	32	6.4
Cromona	0	0.0
Cumarina	3	0.6
Ditolano	0	0.0
Esteroides	24	4.8
Flavonoides	55	11.0
Glicerofosfocolina	0	0.0
Glicerofosfoetanolamina	0	0.0
Lisofosfolípido	1	0.2
Lignan	12	2.4
Nucleósido	1	0.2
Nucleótido	1	0.2
Poliketido	1	0.2
Polipirrol	1	0.2
Porfirina	1	0.2
Terpenoides	56	11.2
Tioester	0	0.0
Tiomorfolina	0	0.0
No identificados	263	52.5
Total de Metabolitos	501	100

Diagrama circular de metabolitos volátiles



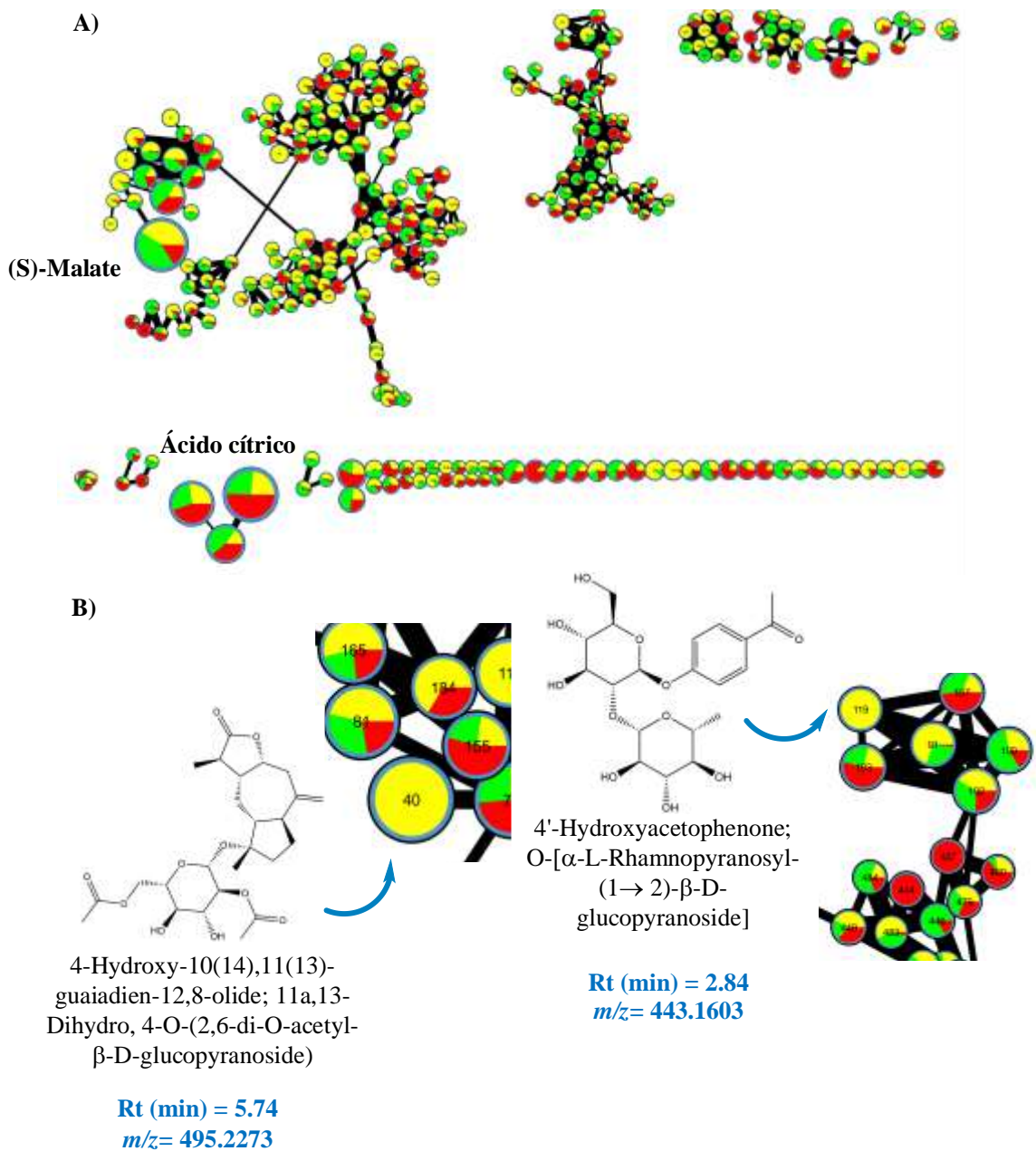


Figura 116. Red molecular de similaridad (Molecular Networking) para ESI (-). En la figura A se muestra la visualización de la red molecular de nodos en gráficos Pie. Cada nodo representa un metabolito y las porciones coloreadas en el gráfico Pie corresponde a las áreas relativas de las tres especies, *C. baccatum* (rojo), *C. chinense* (verde) y *C. frutescens* (amarillo). La red molecular fue construida utilizando el software MetGem. El volumen del nodo indica el valor de importancia (VIP). En B, son dos cortes ampliados donde se indican dos metabolitos biomarcadores identificados para la especie *C. frutescens* (ide 40 y 119) y dos metabolitos biomarcadores con identificación (id) “444” desconocido y “437” con nombre iupac: 22,23-Epoxy-24-norcholest-9(11)-ene-3,6,20-triol; 3-Sulfate,6-O-[D-xylopyranosyl-(1→3)-6-deoxy-β-D-galactopyranosyl-(1→2)-6-deoxy-β-D-lucopyranosyl-(1→4)-[6-deoxy-β-D-glucopyranosyl-(1→2)]-β-D-xylopyranosyl-(1→3)-6-deoxy-β-D-glucopyranoside]) para la especie *C. baccatum*.

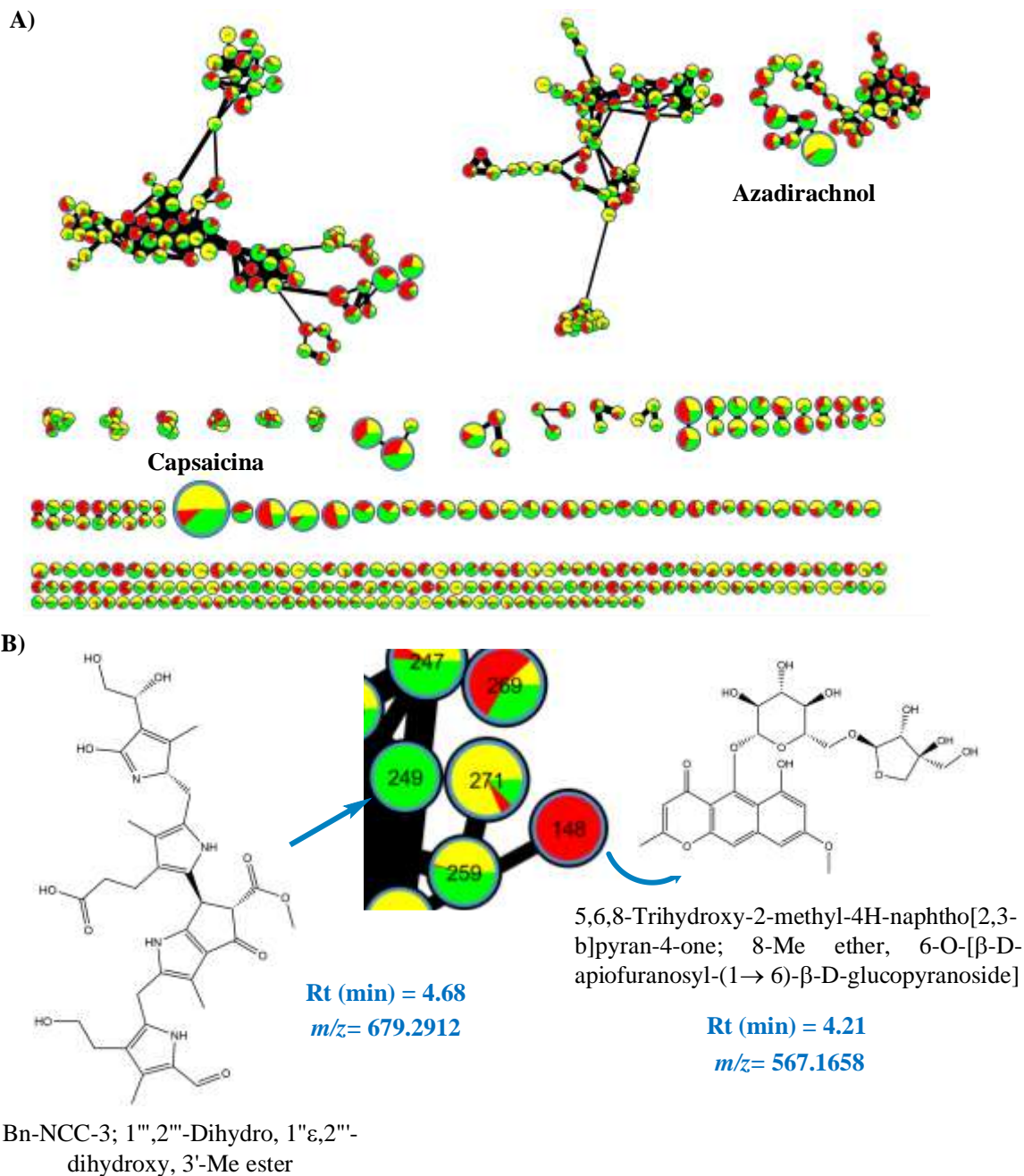


Figura 117. Red molecular de similaridad (Molecular Networking) para ESI (+). En la figura A se muestra la visualización de la red molecular con nodos en gráficos Pie. En la figura B, se observa un corte de la red molecular y se indican dos metabolitos biomarcadores identificados para la especie *C. chinense* con id “249” y *C. baccatum* con ide “148”.

7.8. Validación del modelo de huella metabólica

Para la validación de la huella metabólica se realizó una validación externa del modelo OPLS-DA de la data cromatográfica de los ajíes de Chincha. La validación del modelo OPLS-DA fue realizada con conjuntos de predicción compuesto de réplicas del experimento en diferentes localidades (Chiclayo y Huaral). Desde un punto de vista práctico, la validez de un modelo necesita ser verificada para predecir un fenómeno dado a partir de una nueva data experimental, de tal manera que la certidumbre de los resultados sea confiable. Hay algunos ejemplos que pueden poner en contexto el significado de lo que puede lograrse a través de una validación. Por ejemplo, en el control de calidad de materiales de Ginseng en la búsqueda de su autenticidad, aplicando espectroscopía vibracional (NIR y MIR) como técnicas analíticas (122). Otra aplicación se reporta en el estudio realizado por Guillaume *et al.* (123), que se enfocó en la predicción del tipo de solvente utilizado en la producción de extractos de *Serenoa repens* de cuatro marcas comerciales, preparadas con tres solventes de extracción: hexano, etanol y CO₂ supercrítico, que se predijeron utilizando un modelo OPLS-DA construido con la data cromatográfica de los extractos de la planta preparados en el laboratorio. Para ello se desarrolló un enfoque metabolómico untargeted con data cromatográfica UHPLC-HRMS. Los extractos preparados en el laboratorio fueron utilizados como la data para construir el modelo OPLS-DA o data de entrenamiento (Training set), mientras que los extractos comerciales fueron utilizados como la data a predecir. Se

logró discriminar los tipos de solventes utilizados, destacar los metabolitos biomarcadores y predecir el tipo de solvente utilizado por las marcas comerciales.

En la validación externa del modelo OPLS-DA de Chíncha, y por ende de los biomarcadores, la data de Chíncha fue utilizada para construir el modelo OPLS-DA de discriminación. A este conjunto se le llama conjunto de entrenamiento (Training set). Después de obtener el modelo OPLS-DA del conjunto de datos de entrenamiento, se estimaron los nuevos scores de los conjuntos de datos de prueba o worksets (Predictsets) de la data cromatográfica de Chiclayo y Huaral (figuras 116 y 117). Las figuras 116B y 117B muestran los resultados de scores del conjunto de datos de prueba del análisis por especies. Las figuras 120B y 121B muestran los resultados de scores del conjunto de datos de prueba del análisis por accesiones. Las figuras 118 y 119 muestran la lista de clasificación en SIMCA 15 que muestra los valores Y predichos o pronosticados (YPredPS) tanto para el conjunto de trabajo (data cromatográfica de Chíncha), como para el conjunto de predicción (data cromatográfica de Chiclayo y Huaral) basados en el modelo OPLS-DA y visualiza las probabilidades del conjunto de predicción en los modelos seleccionados (Chiclayo y Huaral) en el análisis por especies. La lista de clasificación para los modelos de análisis discriminante proporciona el valor Y pronosticado para las variables ficticias (0 o 1) utilizadas para dirigir la proyección y representa las observaciones del conjunto de predicciones como filas, y la variable ficticia original en YVarPS y la variable ficticia pronosticada en YPredPS, como columnas. YpredPS son valores predichos e indican la probabilidad de predicción de

pertenencia de las clases (especies y accesiones) de los modelos OPLD-DA de Chiclayo y Huaral al modelo OPLS-DA de Chincha. Los valores <0.35 son de color blanco (indica la no pertenencia a la clase), valores entre 0,35 y 0,65 son de color naranja (están en el límite de la clase) y por encima de 0.65 son verdes (indican la pertenencia a la clase). Las probabilidades >0.65 se encuentran dentro del 90% del nivel de confianza en la curva de probabilidad normal. Las tablas 43-46 muestran los resultados de YPredPS. Claramente se observa que las accesiones y especies fueron predichas ($YPredPS > 0.65$). Las figuras 122 y 123 muestran los gráficos de distancia D_{modXPS} del análisis de predicciones de los modelos OPLS-DA predichos por accesiones y especies. Estos gráficos permiten considerar valores atípicos moderados en las muestras, se observa un solo valor atípico en el análisis por accesiones y especies. La figura 124 muestra los diagramas de barras de los promedios de probabilidades de las clases predichas para las accesiones y especies que muestran una alta probabilidad (> 0.5). Los resultados permiten validar la huella metabólica de las accesiones y especies (biomarcadores); mantienen una independencia de las localidades donde fueron cultivados los ajíes, y permiten la identificación inequívoca de la accesión o especie de ají.

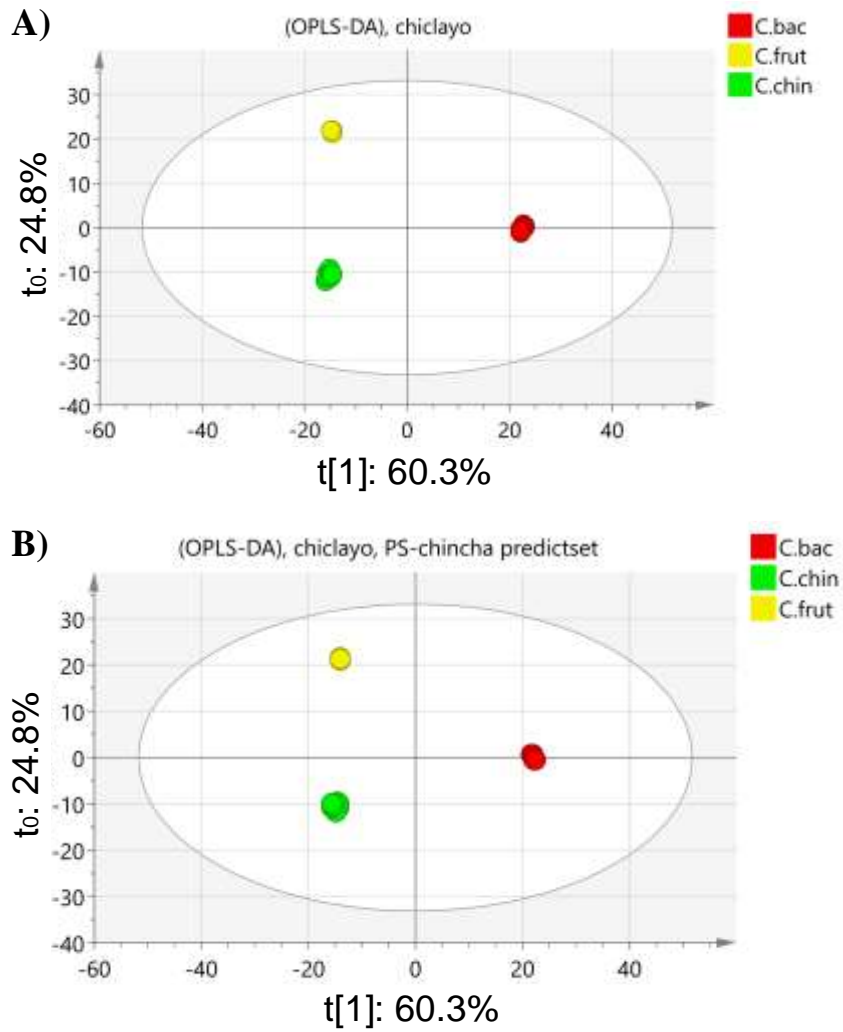


Figura 116. Diagramas de scores del modelo OPLS-DA de la data de Chiclayo (A), CV-ANOVA ($p = 0$), y scores OPLS-DA predicho (B) a partir del modelo OPLS-DA de la data de Chincha.

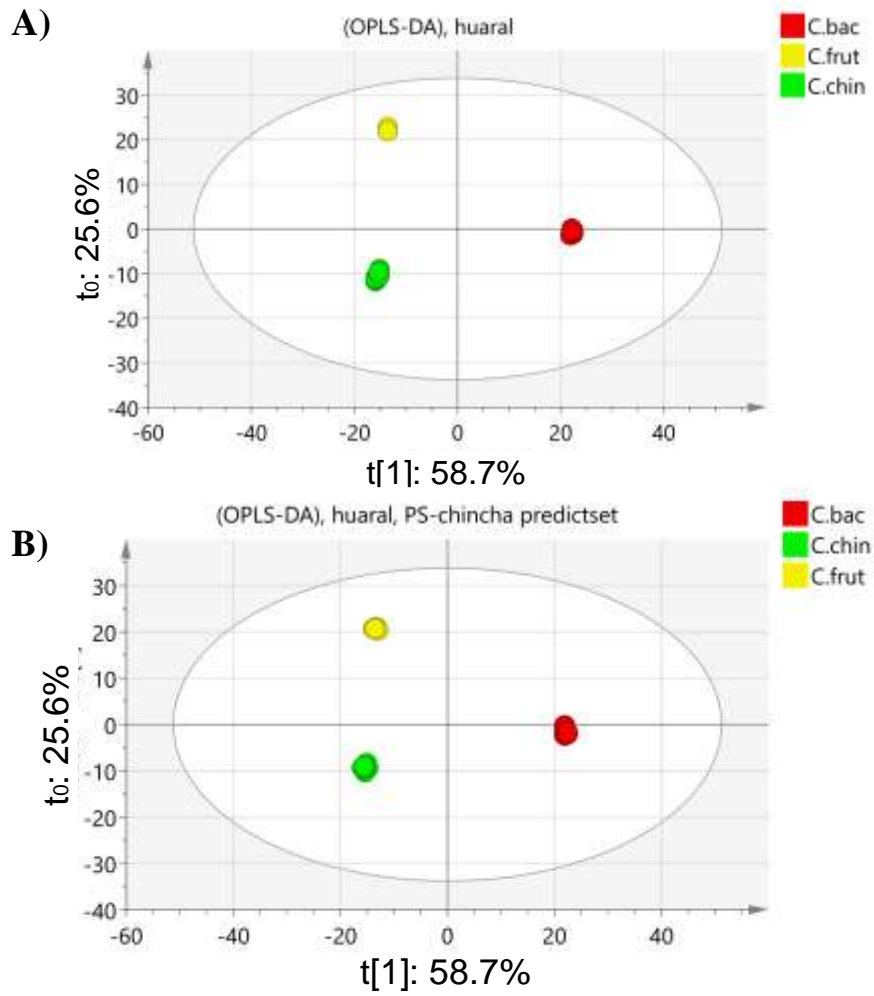


Figura 117. Diagramas de scores del modelo OPLS-DA de la data de Huaral (A) y scores OPLS-DA predicho (B) a partir del modelo OPLS-DA de la data de Chincha.

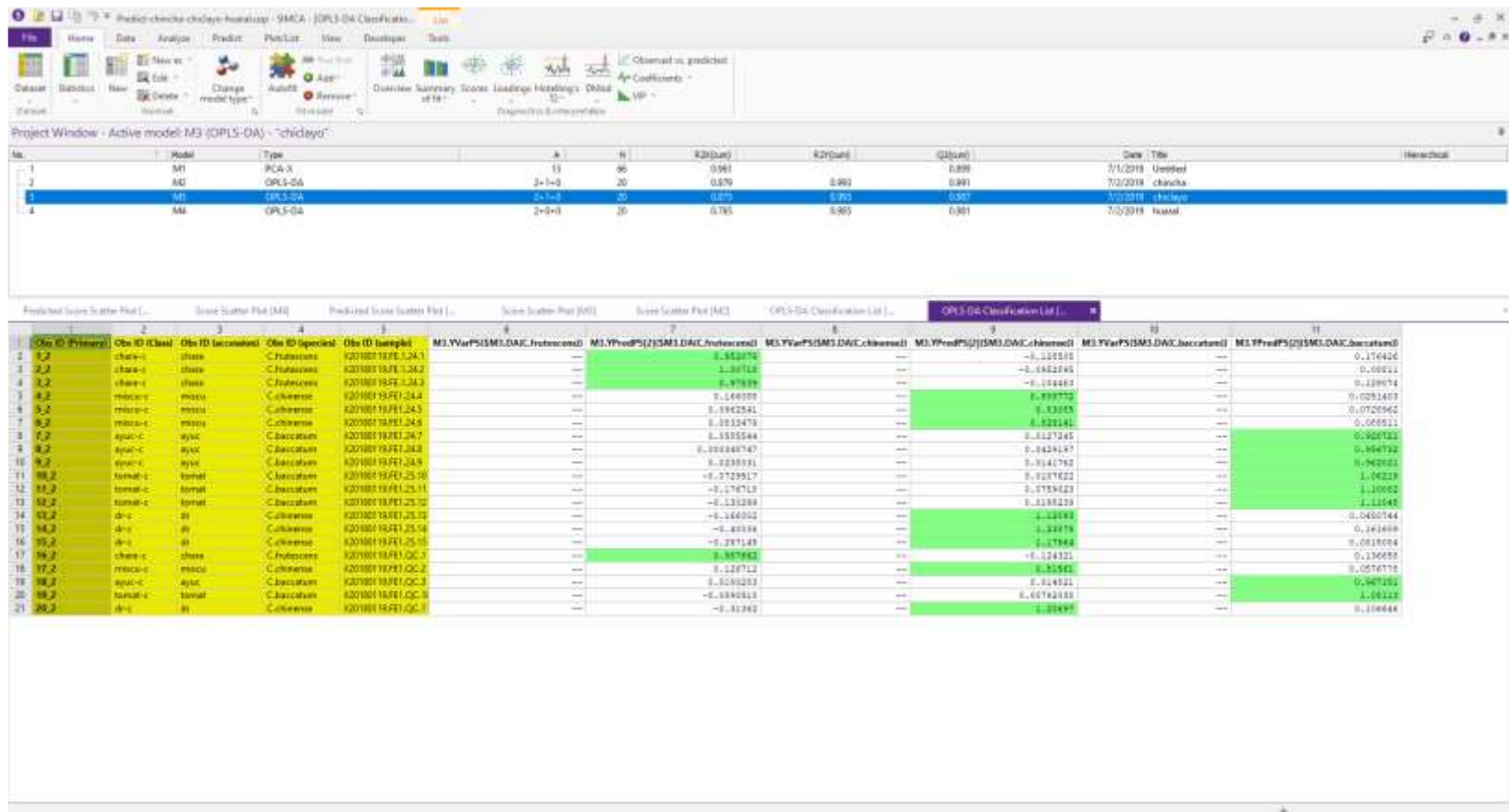


Figura 118. Lista de clasificación de las observaciones del conjunto de predicciones del modelo OPLS-DA de Chiclayo. Provee los valores de probabilidad YpredPS, la probabilidad de ajustar la data de Chiclayo al modelo OPLD-DA de Chianza.

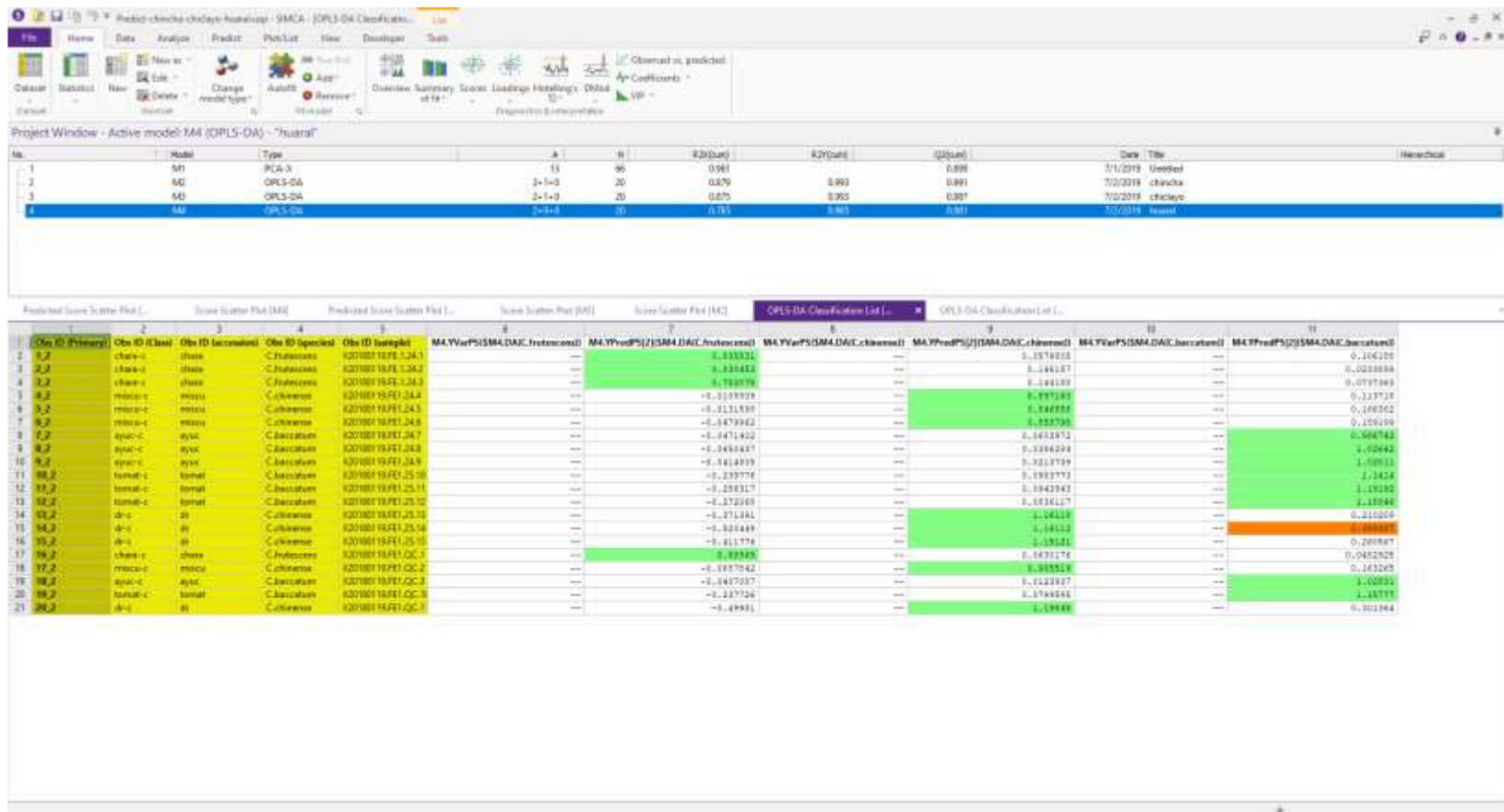


Figura 119. Lista de clasificación de las observaciones del conjunto de predicciones del modelo OPLS-DA de Huaral. Provee los valores de probabilidad YpredPS, la probabilidad de ajustar la data de Huaral al modelo OPLD-DA de Chincha.

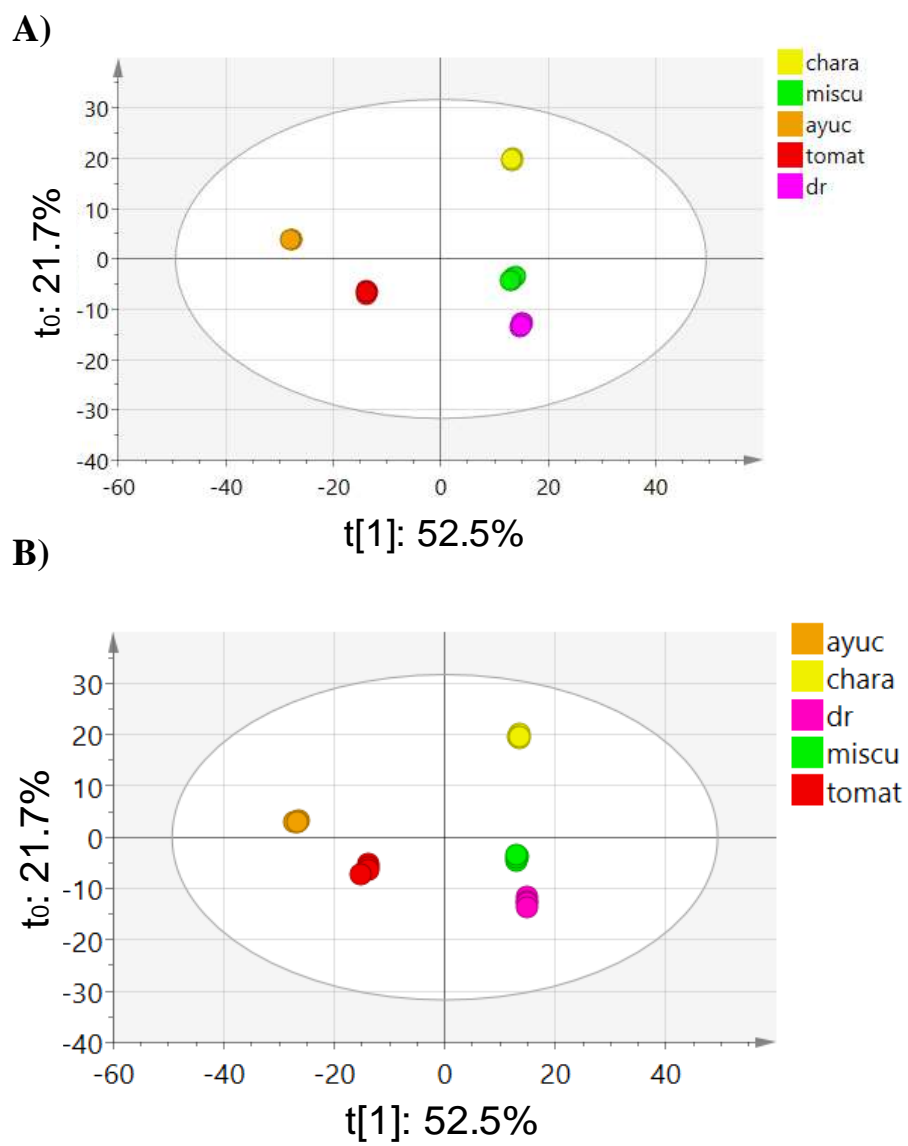


Figura 120. Diagramas de scores del modelo OPLS-DA de la data de Chiclayo (A) y scores OPLS-DA predicho (B) a partir del modelo OPLS-DA de la data de Chíncha.

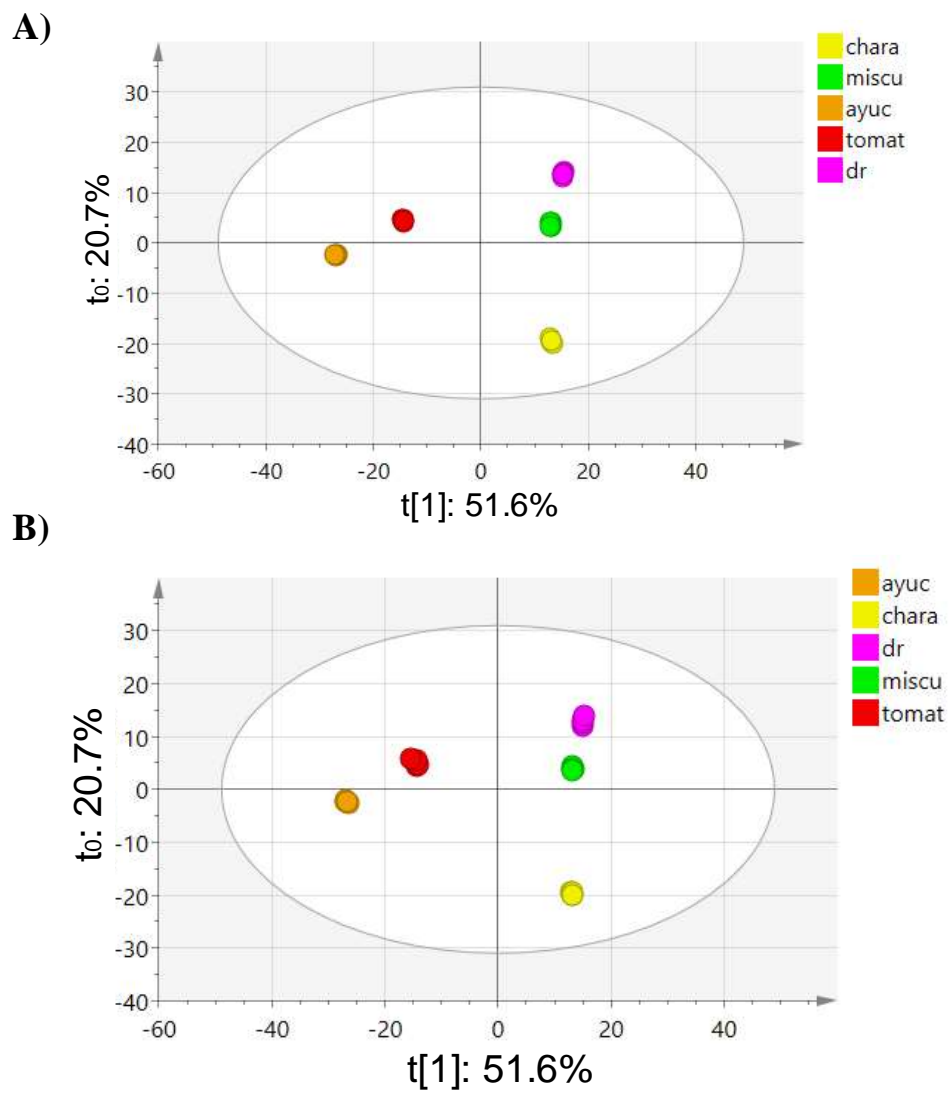


Figura 121. Diagramas de scores del modelo OPLS-DA de la data de Huaral (A) y scores OPLS-DA predicho (B) a partir del modelo OPLS-DA de la data de Chíncha.

Tabla 43. Lista de clasificación que muestra los valores Y pronosticados (YPredPS) para el conjunto de predicción (data de Chiclayo) basado en el modelo OPLS-DA.

Workset (WS)	especies	YPredPS (<i>C. frutescens</i>)	YPredPS (<i>C. chinense</i>)	YPredPS (<i>C. baccatum</i>)
chara-c	<i>C. frutescens</i>	0.996517	-0.0129582	0.0164409
chara-c	<i>C. frutescens</i>	0.976934	0.00612316	0.0169427
chara-c	<i>C. frutescens</i>	0.976591	0.0105165	0.0128926
chara-c	<i>C. frutescens</i>	0.980924	0.00735903	0.0117165
miscu-c	<i>C. chinense</i>	0.0192987	0.989831	-0.00913018
miscu-c	<i>C. chinense</i>	0.00145946	1.01081	-0.012273
miscu-c	<i>C. chinense</i>	0.00903779	0.99291	-0.00194785
miscu-c	<i>C. chinense</i>	0.0232915	0.988145	-0.0114365
dr-c	<i>C. chinense</i>	0.0171089	0.952572	0.0171089
dr-c	<i>C. chinense</i>	0.022315	0.966131	0.022315
dr-c	<i>C. chinense</i>	0.0258636	0.984128	0.0258636
dr-c	<i>C. chinense</i>	0.0124682	1.02075	0.0124682
ayuc-c	<i>C. baccatum</i>	0.0352901	-0.0110603	0.97577
ayuc-c	<i>C. baccatum</i>	0.0283612	-0.00957	0.981209
ayuc-c	<i>C. baccatum</i>	0.0135549	-0.00072825	0.987173
ayuc-c	<i>C. baccatum</i>	0.022328	-0.00332722	0.980999
tomat-c	<i>C. baccatum</i>	0.031341	-0.010155	0.978814
tomat-c	<i>C. baccatum</i>	0.0133711	0.00252476	0.984104
tomat-c	<i>C. baccatum</i>	-0.00751488	0.0198666	0.987648
tomat-c	<i>C. baccatum</i>	-0.0121795	0.0135707	0.998609

Tabla 44. Lista de clasificación que muestra los valores Y pronosticados (YPredPS) para el conjunto de predicción (data de Huaral) basado en el modelo OPLS-DA.

Workset (WS)	especies	YPredPS (<i>C. frutescens</i>)	YPredPS (<i>C. chinense</i>)	YPredPS (<i>C. baccatum</i>)
chara-c	<i>C. frutescens</i>	0.960454	0.0305605	0.00898528
chara-c	<i>C. frutescens</i>	0.938851	0.0481538	0.0129954
chara-c	<i>C. frutescens</i>	0.948214	0.0536885	-0.00190228
chara-c	<i>C. frutescens</i>	0.951446	0.0487604	-0.000206649
miscu-c	<i>C. chinense</i>	0.0432294	0.967548	-0.0107773
miscu-c	<i>C. chinense</i>	0.0304359	0.973318	-0.00375378
miscu-c	<i>C. chinense</i>	0.0345967	0.961983	0.00342059
miscu-c	<i>C. chinense</i>	0.0478237	0.955426	-0.00324991
dr-c	<i>C. chinense</i>	0.0662789	0.927111	0.00660986
dr-c	<i>C. chinense</i>	0.0617709	0.929636	0.00859278
dr-c	<i>C. chinense</i>	0.0304461	0.951857	0.0176967
dr-c	<i>C. chinense</i>	0.00969955	0.987939	0.00236183
ayuc-c	<i>C. baccatum</i>	-0.0250753	0.0305605	0.992284
ayuc-c	<i>C. baccatum</i>	-0.0473992	0.0481538	1.00313
ayuc-c	<i>C. baccatum</i>	-0.0539996	0.0536885	0.994982
ayuc-c	<i>C. baccatum</i>	-0.0463374	0.0487604	0.996214
tomat-c	<i>C. baccatum</i>	0.0115917	0.00102434	0.987384
tomat-c	<i>C. baccatum</i>	0.00766845	0.0030055	0.989326
tomat-c	<i>C. baccatum</i>	-0.0233678	0.0249865	0.998381
tomat-c	<i>C. baccatum</i>	-0.0363732	0.0338236	1.00255

Tabla 45. Lista de clasificación que muestra los valores Y pronosticados (YPredPS) para el conjunto de predicción (data de Chiclayo) basado en el modelo OPLS-DA.

Workset (WS)	especies	YPredPS (Charapita)	YPredPS (Miscucho)	YPredPS (Ayuclo)	YPredPS (Tomatito rojo)	YPredPS (Dulce rojo)
Chara-c	<i>C. frutescens</i>	1.01484	0.024239	-0.0342364	0.0386956	-0.0435426
Chara-c	<i>C. frutescens</i>	0.994182	-0.00283867	-0.0012919	0.000633702	0.00931504
Chara-c	<i>C. frutescens</i>	0.990934	0.00837842	-0.0186134	0.0210585	-0.00175758
Chara-c	<i>C. frutescens</i>	0.996562	0.0347972	-0.0285279	0.029413	-0.0322447
Miscu-c	<i>C. chinense</i>	0.0144328	0.950389	0.0129904	-0.0102661	0.0324541
Miscu-c	<i>C. chinense</i>	-0.0172153	0.983529	0.001577	0.0123313	0.0197778
Miscu-c	<i>C. chinense</i>	-0.0132196	1.02033	0.00700091	-0.000761375	-0.0133476
Miscu-c	<i>C. chinense</i>	0.00570413	0.997643	0.0318563	-0.0371121	0.00190812
Ayuc-c	<i>C. baccatum</i>	0.0153707	0.00723267	0.913587	0.0797713	-0.0159617
Ayuc-c	<i>C. baccatum</i>	-0.0111972	0.0485235	0.916949	0.0829418	-0.0372171
Ayuc-c	<i>C. baccatum</i>	-0.012313	0.0171126	0.936953	0.071787	-0.0135397
Ayuc-c	<i>C. baccatum</i>	0.00312553	0.00236675	0.919749	0.0818103	-0.00705136
Tomat-c	<i>C. baccatum</i>	0.0366984	-0.0274033	0.0550812	0.919151	0.0164722
Tomat-c	<i>C. baccatum</i>	0.00524361	0.0198011	0.0383035	0.943933	-0.00728081
Tomat-c	<i>C. baccatum</i>	0.00552857	-0.0367484	0.0495927	0.933856	0.0477706
Tomat-c	<i>C. baccatum</i>	-0.0122086	-0.173019	0.1237	0.873518	0.188011
Dr-c	<i>C. chinense</i>	0.0425078	-0.00823259	0.020119	-0.0318071	0.977413
Dr-c	<i>C. chinense</i>	0.0102618	0.0399114	0.000479698	-0.00317839	0.952525
Dr-c	<i>C. chinense</i>	0.00990544	-0.0159729	0.00949712	-0.010122	1.00669
Dr-c	<i>C. chinense</i>	-0.0349242	0.0367981	0.0182328	-0.0249557	1.00485

Tabla 46. Lista de clasificación que muestra los valores Y pronosticados (YPredPS) para el conjunto de predicción (data de Huaral) basado en el modelo OPLS-DA.

Workset (WS)	Especies	YPredPS (Charapita)	YPredPS (Miscucho)	YPredPS (Ayucillo)	YPredPS (Tomatito rojo)	YPredPS (Dulce rojo)
Chara-c	<i>C. frutescens</i>	1.01068	0.0454083	-0.0573309	0.0747671	-0.0735203
Chara-c	<i>C. frutescens</i>	0.990515	0.0234584	-0.0285215	0.0475741	-0.0330259
Chara-c	<i>C. frutescens</i>	0.994454	0.0308726	-0.0290076	0.0361295	-0.032449
Chara-c	<i>C. frutescens</i>	0.998832	0.0569835	-0.00426875	-0.000786617	-0.05076
Miscu-c	<i>C. chinense</i>	0.0223874	0.917941	0.0167603	-0.0312482	0.07416
Miscu-c	<i>C. chinense</i>	-0.0062516	0.951443	0.00608061	-0.009267	0.0579946
Miscu-c	<i>C. chinense</i>	-0.000779942	0.985924	0.00664055	-0.0120383	0.0202535
Miscu-c	<i>C. chinense</i>	0.0121051	0.976633	0.0282328	-0.039161	0.0221904
Ayuc-c	<i>C. baccatum</i>	0.0286976	-0.0541133	0.939645	0.0718694	0.0139017
Ayuc-c	<i>C. baccatum</i>	0.00123639	0.0103632	0.978693	0.0269108	-0.0172036
Ayuc-c	<i>C. baccatum</i>	0.00831182	-0.0602887	1.00216	0.000562653	0.0492575
Ayuc-c	<i>C. baccatum</i>	0.0114489	-0.0393683	0.977942	0.0281487	0.021829
Tomat-c	<i>C. baccatum</i>	0.00651085	-0.0291217	0.0783123	0.883763	0.0605351
Tomat-c	<i>C. baccatum</i>	-0.0127079	0.0106202	0.0807417	0.876765	0.044581
Tomat-c	<i>C. baccatum</i>	-0.0258646	-0.032608	0.0656955	0.909612	0.0831653
Tomat-c	<i>C. baccatum</i>	-0.0305769	-0.139111	0.181977	0.78719	0.200521
Dr-c	<i>C. chinense</i>	0.0223173	0.0953796	-0.0422523	0.060197	0.864358
Dr-c	<i>C. chinense</i>	0.00229232	0.136213	-0.0403555	0.053786	0.848065
Dr-c	<i>C. chinense</i>	-0.0114989	0.0937748	-0.0559268	0.0873013	0.886349
Dr-c	<i>C. chinense</i>	-0.040172	0.131569	-0.027517	0.0360304	0.90009

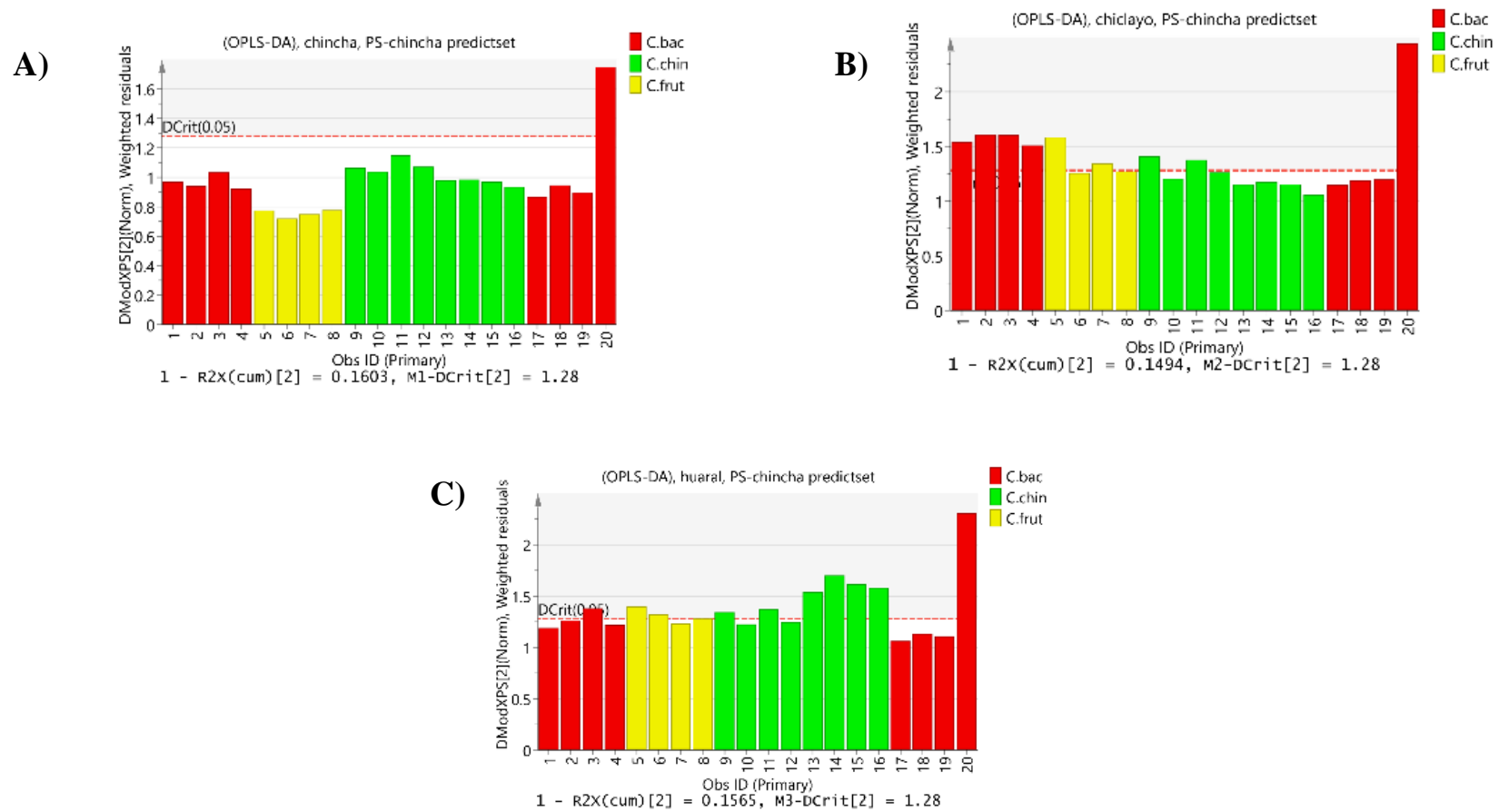


Figura 122. Gráficos DModX para la detección de valores atípicos moderados y fuertes predichos, por especies en **A)** DmodXPS de la data de Chincha, data de entrenamiento, en **B)** DmodXPS de la data predicha de Chiclayo (Workset1), en **C)** DmodXPS de la data predicha de Huaral (Workset2), Dcritico = 0.05 para un nivel de significancia del 95%.

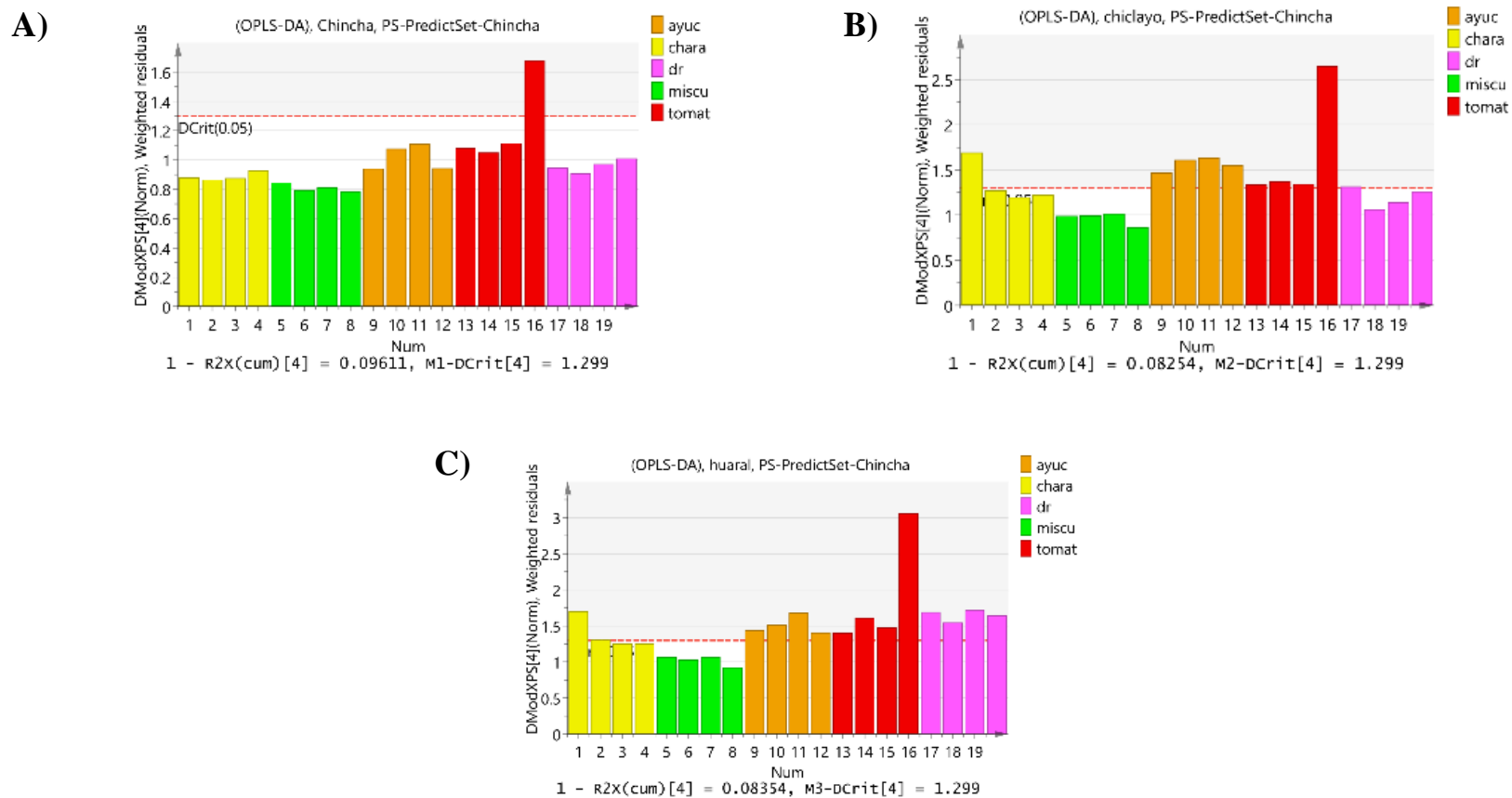


Figura 123. Gráficos DModX para la detección de valores atípicos moderados y fuertes predichos, por accesiones en **A)** DmodXPS de la data de Chincha, data de entrenamiento, en **B)** DmodXPS de la data predicha de Chiclayo (Workset1), en **C)** DmodXPS de la data predicha de Huaral (Workset2), Dcritico = 0.05 para un nivel de significancia del 95%.

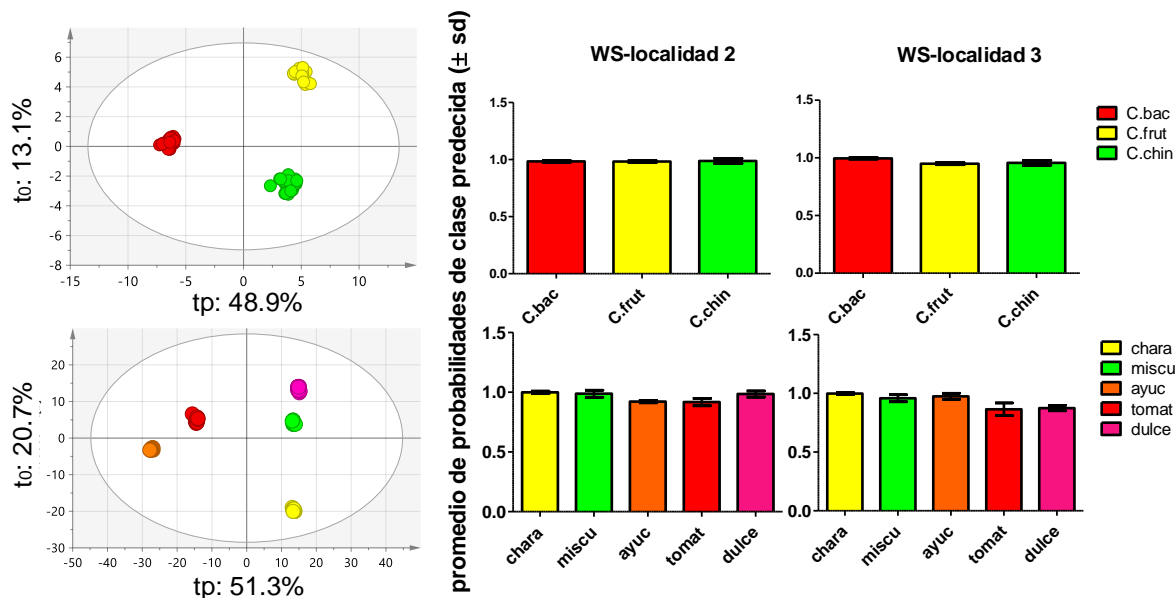


Figura 124. Gráficos de OPLSDA y probabilidades de clases predicha. Se observa el OPLS-DA por especie y accesión de las tres localidades con una varianza explicada de 62% y 72%, respectivamente. La data cruda se separó en tres matrices de datos de 20 individuos cada uno con el mismo número de variables (metabolitos) en función de la localidad 1, 2 y 3. A continuación, con la data de la localidad 1, se construyó un modelo de OPLS-DA, por especie y accesión que sirvió como conjunto de entrenamiento (Workset). La data de la localidad 2 y 3 sirvieron como conjunto de prueba. Los resultados mostrados en el gráfico de barras muestran las probabilidades de clase predichas de la clasificación OPLS-DA. Un valor superior a 0.5 indica un alto nivel de confianza de la clase predicha. Aquí: Charapita (chara, *C. frutescens* = C. frut), Miscucho (miscu, *C. chinense* = C. chin), Ayucllo (ayuc, *C. baccatum* = C.bac), Tomatito rojo (tomat, *C. baccatum*) y Dulce rojo (dulce, *C. chinense*), (sd = desviación estándar).

7.9. Identificación de descriptores químicos del sabor y aroma

7.9.1. Análisis multibloque

La data cromatográfica obtenida como áreas relativas de los picos cromatográficos de metabolitos (volátiles y no volátiles) se integraron con la data sensorial (score del sabor y aroma) para encontrar las correlaciones en estos conjuntos de datos. El olor y sabor están relacionados con los metabolitos volátiles y no volátiles, respectivamente. También, la integración de los bloques de datos tiene en cuenta la similitud en los patrones de expresión del conjunto de muestras y por lo tanto están relacionados. Las dos plataformas analíticas (UHPLC y GC-MS) y “plataforma sensorial” emplearon estrategias independientes de adquisición de data cuyos resultados son complementarios en un contexto de análisis global del problema, ya que el objetivo fue encontrar los metabolitos que expliquen el sabor y aroma. El AMV de los tres bloques de datos mostraron en la clasificación categórica y discriminante de los modelos estadísticos multivariados de OPLS-DA similares anclajes subyacentes de las correspondencias por accesiones o especies en los puntos de datos, clusterizados, lo que contribuye a una disminución de la posibilidad de falsos descubrimientos en el análisis multibloque. Específicamente, en la estrategia de integración de datos se utilizó un novedoso enfoque que utiliza métodos multivariados especializados capaces de identificar características relevantes de conjuntos de datos de alta dimensionalidad, análisis discriminante de correlación canónica generalizada sparse vía análisis de integración de datos para

el descubrimiento de biomarcadores utilizando componente latentes, estructura que es parte del Package MixOmics del software estadístico de R (68).

Denotamos como X_1 ($N \times P_1$), X_2 ($N \times P_2$) y X_3 ($N \times P_3$) las tres matrices de datos (metabolitos volátiles, metabolitos no volátiles y descriptores del sabor y aroma), donde N es el número total de muestras (en la matriz de datos como filas= accesiones o especies por cuatro réplicas) en los bloques de datos, y por P_1 , P_2 y P_3 las tres variables respuesta o predictores (áreas relativas de metabolitos volátiles, áreas de metabolitos no volátiles y descriptores del sabor y aroma, en columnas de cada bloque de datos, con $N \ll P_1, P_2$ y P_3). El resultado categórico Y (por ejemplo *C. baccatum* vs *C. chinense* vs *C. frutescens*) se expresa como una matriz indicadora ficticia Y (con valores categóricos 0 y 1), donde cada columna representa una categoría de resultado y cada fila indica la pertenencia a la clase de cada muestra. Por lo tanto, Y es de tamaño N observaciones (filas) \times resultado de 3 categorías (columnas). Dichos conjuntos de datos estructurados pueden considerarse como bloques Q de variables, X_1, \dots, X_Q , donde cada bloque X_Q (con $Q = 1, \dots, Q$) representa un conjunto de variables P_Q observadas en el mismo grupo de N individuos (figura 125). La integración N combina diferentes conjuntos de datos ómicos medidos en las mismas N muestras o individuos para clasificar un resultado discreto y seleccionar características de cada conjunto de datos, es decir N -integración con análisis discriminante sparse. El método se basa en parte en el análisis de correlación canónica generalizada (68).

El algoritmo utilizado para realizar el análisis discriminante de mínimos cuadrados parciales sparce (sPLS-DA) es una extensión de la regresión de mínimos cuadrados parciales sparce (sPLS) aplicada a problemas de clasificación. De manera similar a PLS-DA, sPLS-DA se basa en el uso de la regresión de PLS para fines discriminantes (124). Esta versión de sparce tienen como objetivo lograr la selección de variables y la reducción de la dimensionalidad simultánea entre los bloques de datos.

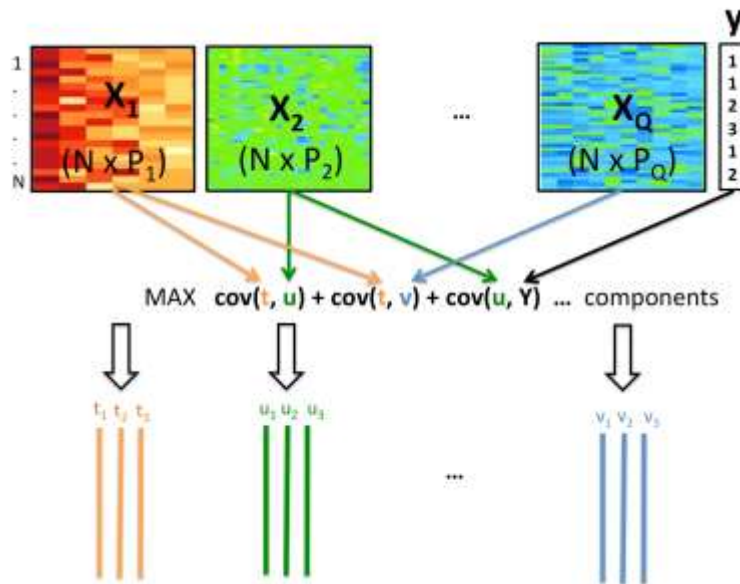


Figura 125. Esquema multibloque utilizando el método S-PLSDA (68).

La selección de variables es un elemento importante ya que contribuye a la construcción de modelos parsimoniosos, es decir, modelos simples con un gran poder predictivo explicativo (125). El problema de la optimización (figura 121) es la maximización de la matriz de covarianzas entre los componentes principales (124) (lograr la máxima correlación entre las variables metabólicas y sensoriales).

Un modelo parsimonioso (sparse) captura la información que permite explicar y predecir eventos naturales. El principio de parsimonia utiliza modelos lo más simples de explicar con un número mínimo de parámetros necesarios (126).

En análisis multivariado el término sparse se refiere a una matriz en la que la mayoría de los elementos son iguales a cero. La idea principal de los métodos sparse es reducir la influencia del ruido contenido en las variables irrelevantes al forzar a los coeficientes del modelo relacionados con esas variables sean iguales a cero y, por consiguiente, realizar una selección de variables (127).

Los resultados del análisis multibloque splsda (MBSPLSDA) utilizando el Package MixOmics se observan en las figuras 126-132. En las figuras 126-129, se observan los resultados del análisis de regresión de mínimos cuadrados parciales sparse (SPLS-DA-R) de las 4 componentes principales.

La regresión de mínimos cuadrados parciales (PLS-R) relaciona el conjunto de variables independientes (predictores) con las variables de respuesta

(observaciones) al reducir la gran cantidad de variables originales a un nuevo espacio de variables basado en una pequeña cantidad de factores ortogonales llamados variables latentes o componentes principales. La información no contenida en los componentes principales se describe mediante un valor residual calculado para cada matriz de datos. El modelo resultante tiene la siguiente forma:

$$Y = Xb + E \dots (1)$$

Donde Y es la matriz de respuestas independientes de mediciones instrumentales y sensoriales (N muestras \times 1), X es la matriz de predictores (N muestras \times Áreas relativas de volátiles, no volátiles o score sensorial), b es la matriz de coeficientes de regresión obtenida del análisis PLS, y E es la matriz de Información residual no contenida en los componentes principales (128).

En las figuras 126-129 se observan el análisis de correlaciones interbloques SPLSDA para las componentes principales de los tres bloques de datos con sus respectivos gráficos multivariados de SPLSDA de cada individuo (4 réplicas por accesión de ají) y sus respectivas elipses de confianza, con un nivel de confianza del 95%. El modelo de regresión SPLSDA de las cinco variables categóricas evalúa las correlaciones de las componentes principales en cada bloque de datos. Para la primera componente (figura 126) se observa el gráfico de correlaciones interbloques, en la que se muestra los coeficientes de regresión de Pearson. Un valor de $R^2 = 1$, denota un modelo SPLSDA de correlación ideal. El coeficiente de

correlación de Pearson entre el bloque de volátiles vs bloque sensorial da un valor de $R^2 = 0.98$, para el bloque No volátiles vs bloque sensorial se observa un valor de $R^2 = 0.91$, y para el bloque volátiles vs bloque No volátiles se observa un $R^2 = 0.96$. Valores altos de R^2 denotan un alto porcentaje de varianza que puede explicar el modelo sparse de mínimos cuadrados parciales, con un bajo porcentaje de residuales (por ejemplo, para $R^2 = 0.98$, significa que la data explica un 98% de la varianza explicada del modelo splsda con un error o residual del 2%). Bajo el mismo criterio, se puede interpretar las correlaciones de la segunda, tercera y cuarta componente principales (figuras 127-129).

La figura 130 muestra el mapa de imagen en clúster (CIM) o Heatmap del modelo SPLSDA para los tres bloques de datos. Las filas y columnas muestran el análisis jerárquico de cluster (HCA) de las asociaciones de accesiones de los 5 ajíes y de los metabolitos volátiles, no volátiles y scores sensoriales respectivamente. Se utilizó la distancia euclidiana y el método Ward para la agrupación jerárquica HCA. En el CIM cada bloque de color representa una asociación entre subconjuntos de las variables X (accesiones) y las variables Y (metabolitos volátiles, no volátiles y score sensorial). En el CIM, cada bloque de color representa una asociación entre subconjuntos de las variables X y las variables Y. El color rojo indica que los grupos X e Y están correlacionados positivamente (grupos de metabolitos volátiles, no volátiles y score sensorial), el color azul indica una correlación negativa entre los grupos de variables X e Y, las variables en verde indican variables con correlaciones débiles mientras que amarillo indica valores de correlación muy débiles (variables

irrelevantes). Los dendrogramas en la parte superior y en el lado izquierdo del mapa indican cómo se unen los grupos, cuanto más larga es la distancia, más nítida es la frontera entre los bloques de colores (129).

La figura 131 muestra la gráfica de individuos de los tres bloques de datos por el método SPLSDA. En cada bloque se observan tendencias similares para la discriminación de las cinco accesiones de ajés.

La gráfica de CircosPlot mostrada en la figura 132 representa las correlaciones entre variables de áreas relativas de metabolitos volátiles, no volátiles y score sensorial. El diagrama de CircosPlot se basa en una matriz de similaridad extendido a múltiples conjuntos de datos. Por defecto, el análisis incluye un argumento de corte cutt-off $r = 0.8$ para visualizar los coeficientes de correlación por encima de este umbral y eliminar algunas asociaciones más débiles. Los resultados del análisis CircosPlot fueron exportados en una plantilla Excel en formato *.CSV. En la tabla Excel exportada se resumieron las asociaciones de metabolitos volátiles, no volátiles con los scores sensoriales (descriptores químicos del sabor y aroma).

Las tablas 47-56 muestran los resultados obtenidos del análisis CircosPlot. Las correlaciones fueron seleccionadas de acuerdo con los resultados del AFM, diagramas Hetmap y VIP del análisis sensorial (figuras 60, 62 y 63).

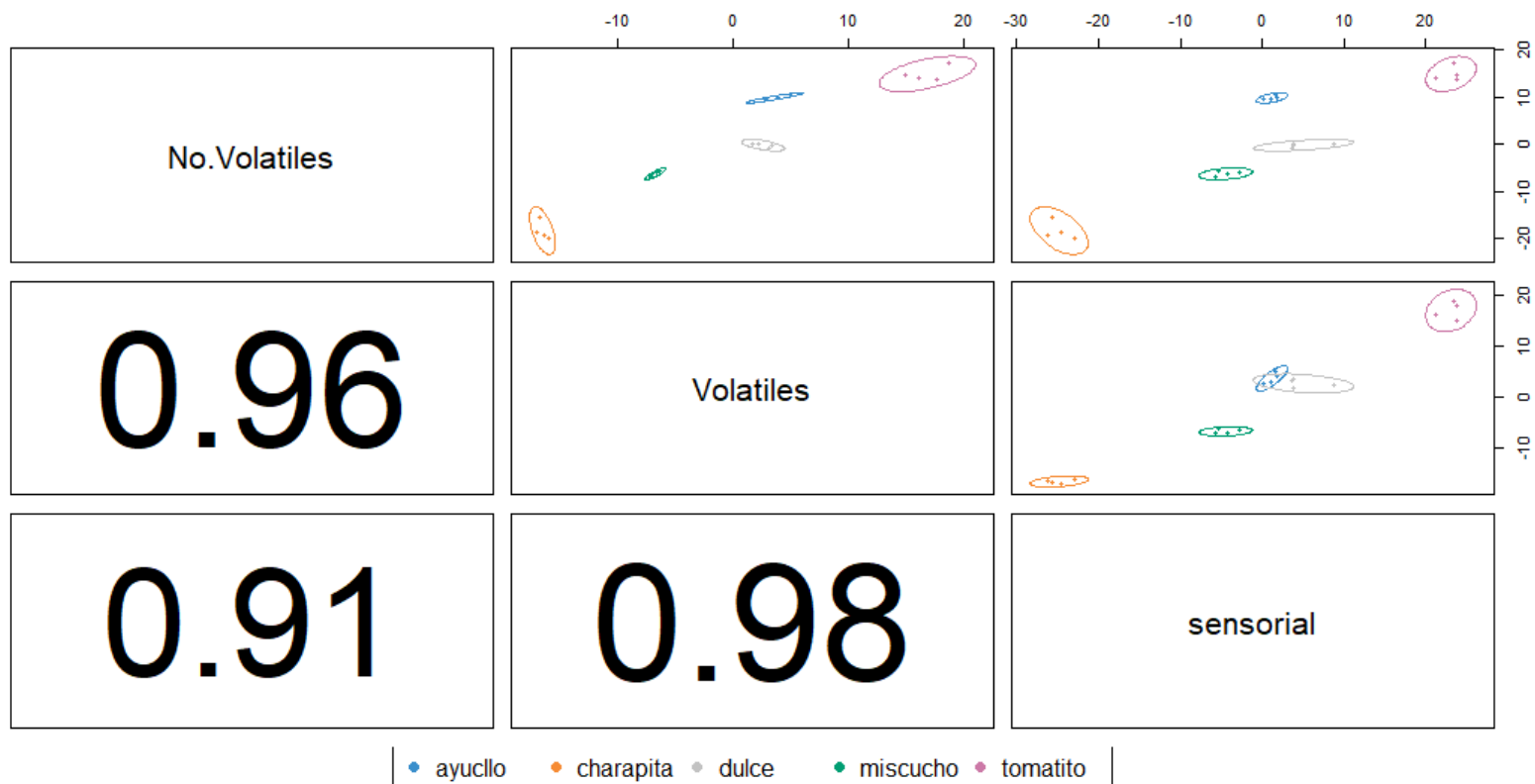


Figura 126. Gráfico de diagnóstico para verificar si la correlación multibloque del modelo S-PLSDA entre los componentes de cada conjunto de datos se ha maximizado. Los primeros componentes de cada conjunto de datos están altamente correlacionados entre sí, siendo la correlación de Pearson $R^2 = 0.91$ para el bloque No-Volátiles vs sensorial, y $R^2 = 0.98$, para el bloque Volátiles vs sensorial. Los colores y elipses indican el poder discriminativo de cada componente para separar las diferentes accesiones de ajés.

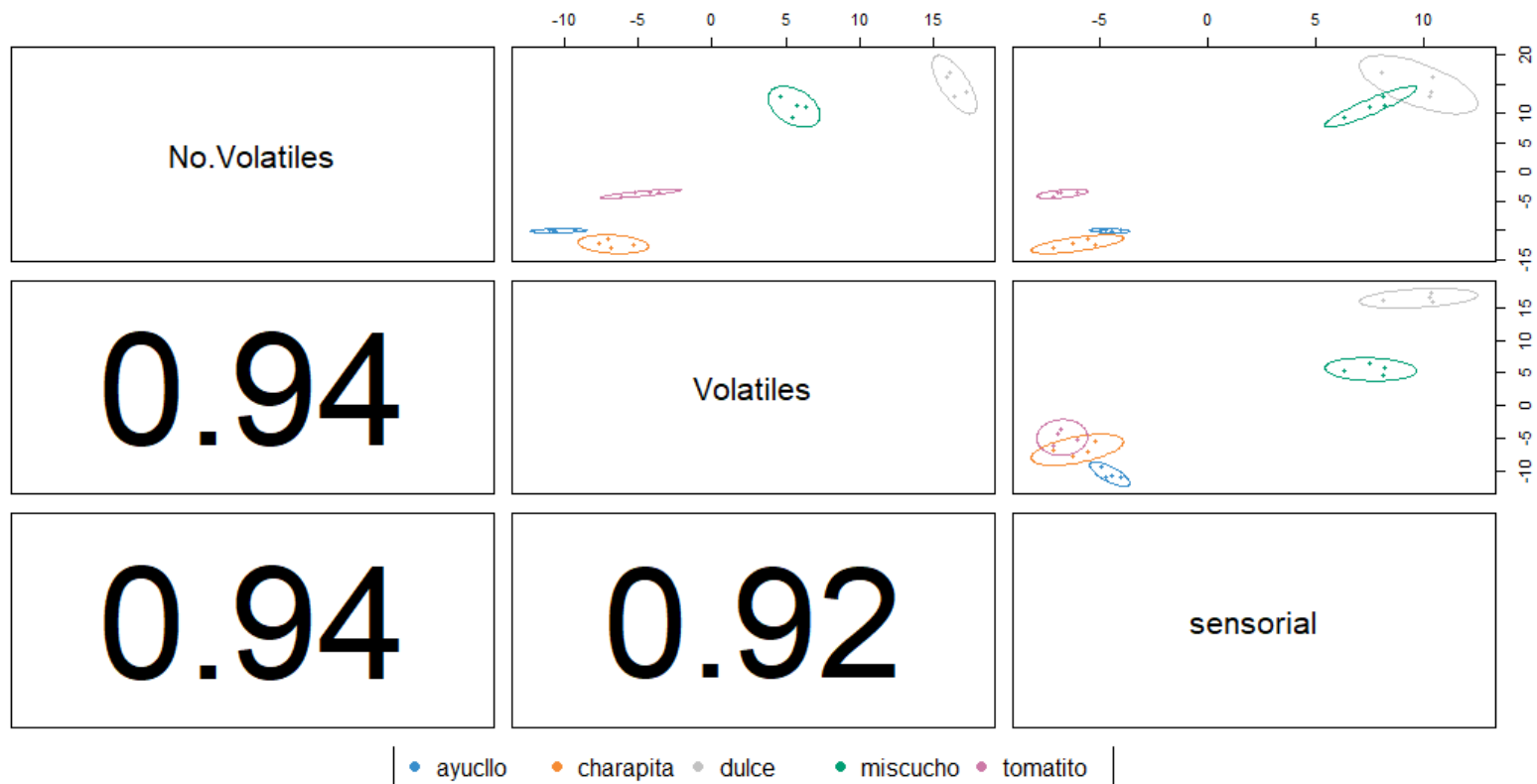


Figura 127. Gráfico de diagnóstico para verificar si la correlación multibloque del modelo S-PLSDA entre los componentes de cada conjunto de datos se ha maximizado. Las segundas componentes de cada conjunto de datos están altamente correlacionadas entre sí, siendo la correlación de Pearson $R^2 = 0.94$ para el bloque No-Volátiles vs sensorial, y $R^2 = 0.92$, para el bloque Volátiles vs sensorial. Los colores y elipses indican el poder discriminativo de cada componente para separar las diferentes accesiones de ajés.

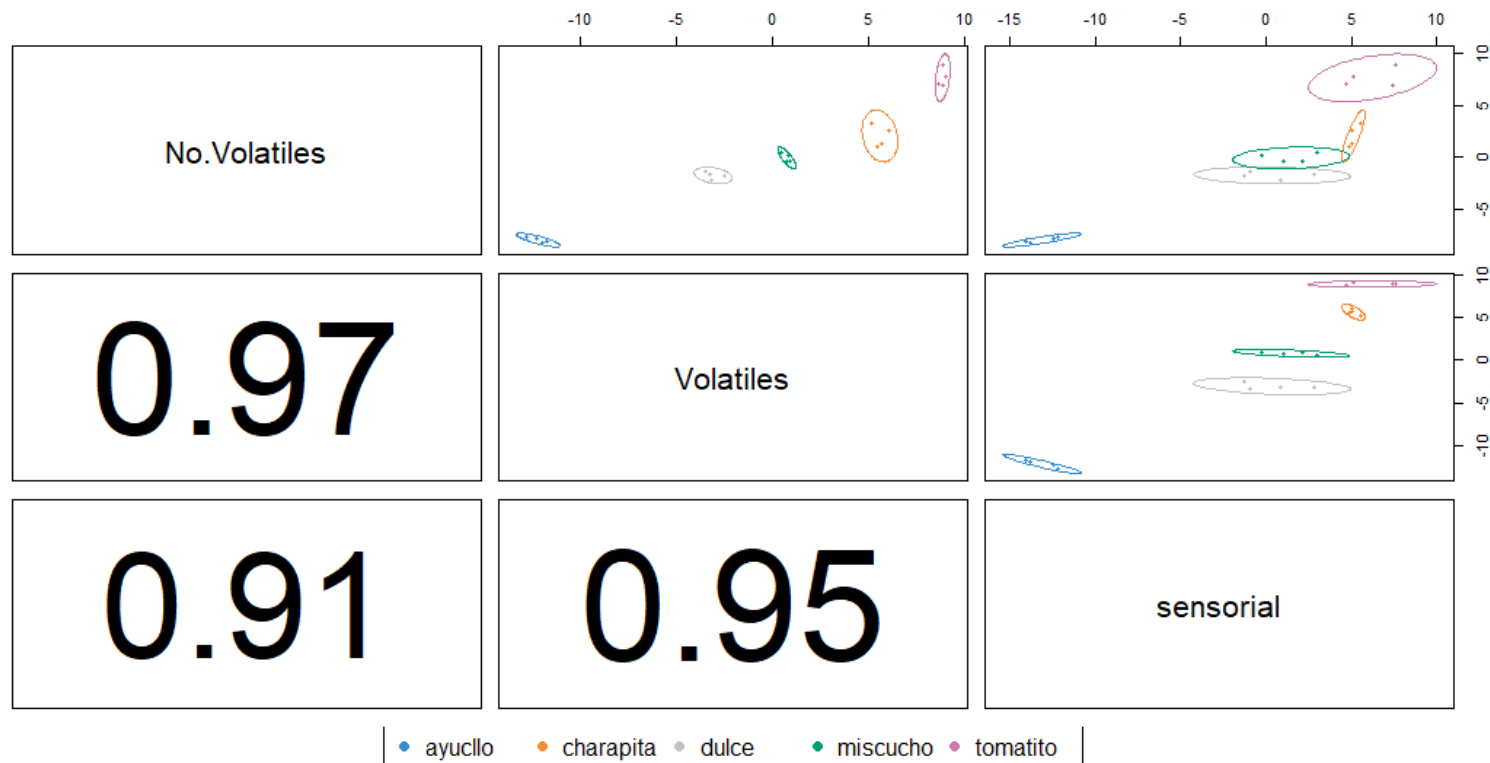


Figura 128. Gráfico de diagnóstico para verificar si la correlación multibloque del modelo S-PLSDA entre los componentes de cada conjunto de datos se ha maximizado. Las terceras componentes de cada conjunto de datos están altamente correlacionadas entre sí, siendo la correlación de Pearson $R^2 = 0.91$ para el bloque No Volátiles vs sensorial, y $R^2 = 0.95$, para el bloque Volátiles vs sensorial. Los colores y elipses indican el poder discriminativo de cada componente para separar las diferentes accesiones de ajíes.

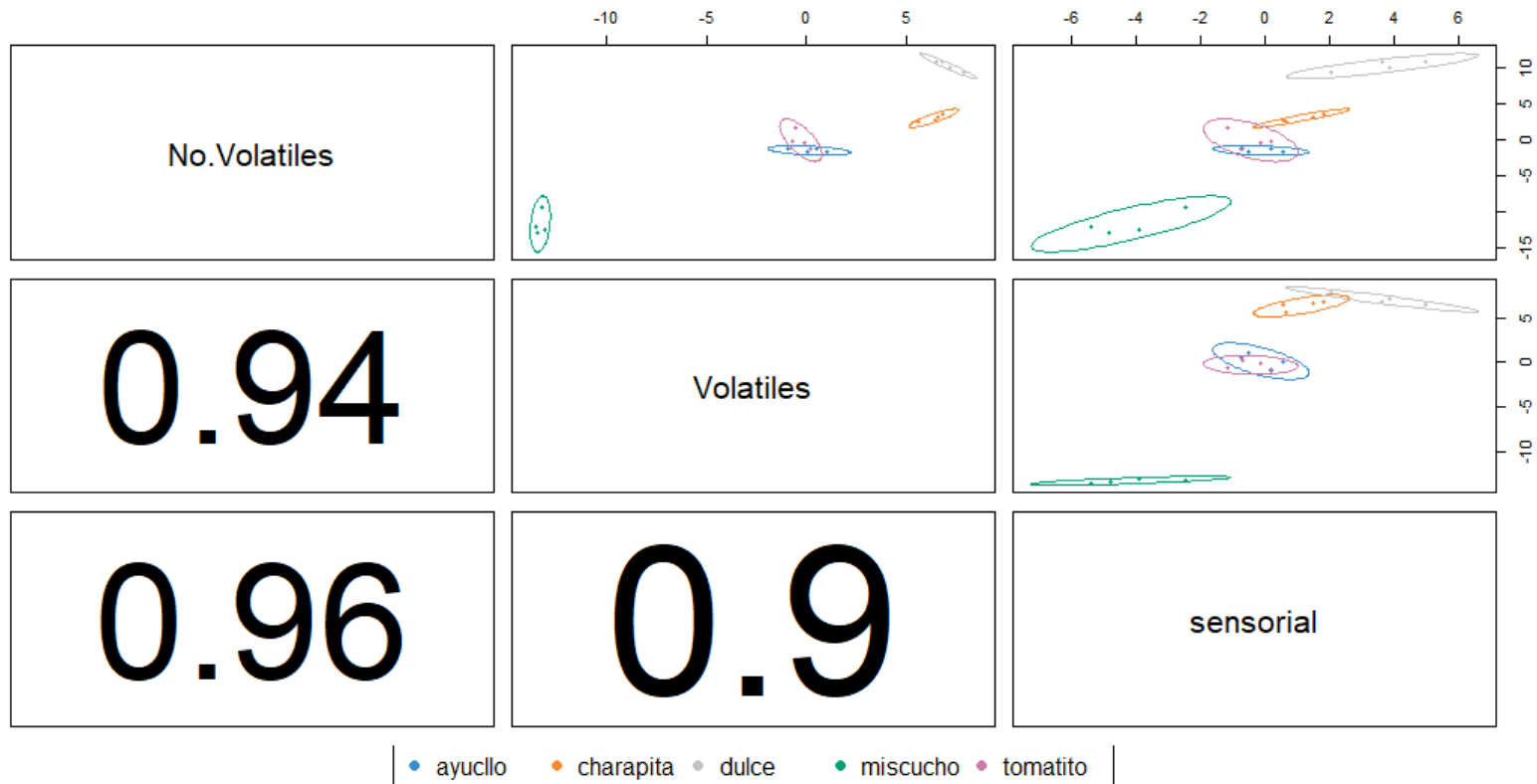


Figura 129. Gráfico de diagnóstico para verificar si la correlación multibloque del modelo S-PLSDA entre los componentes de cada conjunto de datos se ha maximizado. Las cuartas componentes de cada conjunto de datos están altamente correlacionadas entre sí, siendo la correlación de Pearson $R^2 = 0.96$ para el bloque No Volátiles vs sensorial, y $R^2 = 0.9$, para el bloque Volátiles vs sensorial. Los colores y elipses indican el poder discriminativo de cada componente para separar las diferentes accesiones de ajés.

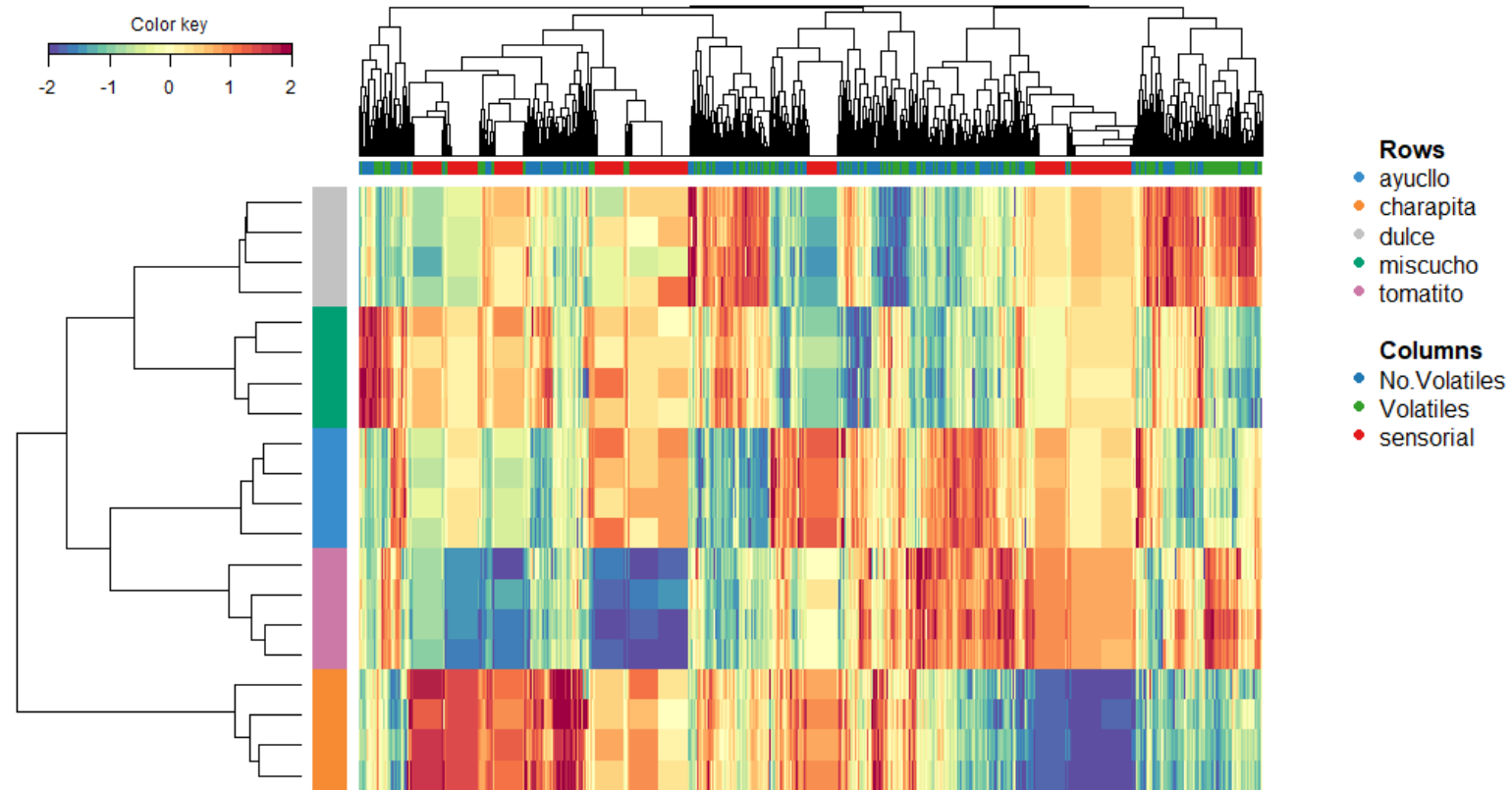


Figura 130. Diagrama CIM (Heatmap) por accesiones del modelo S-PLSDA de las tres tablas de datos. La gráfica muestra las agrupaciones del análisis jerárquico de cluster HCA. El HCA fue construido utilizando la distancia euclidiana y el algoritmo Ward. En las columnas y filas se observa el arreglo jerárquico de los metabolitos volátiles, no volátiles, data sensorial y arreglo categórico de las accesiones, respectivamente. De acuerdo a la escala de colores, las variables en rojo corresponden a correlaciones positivas, las variables en azul a correlaciones negativas, las variables en amarillo son irrelevantes (correlaciones muy débiles). Los bloques de datos de igual color permiten visualizar la discriminación entre las 5 accesiones de ajés.

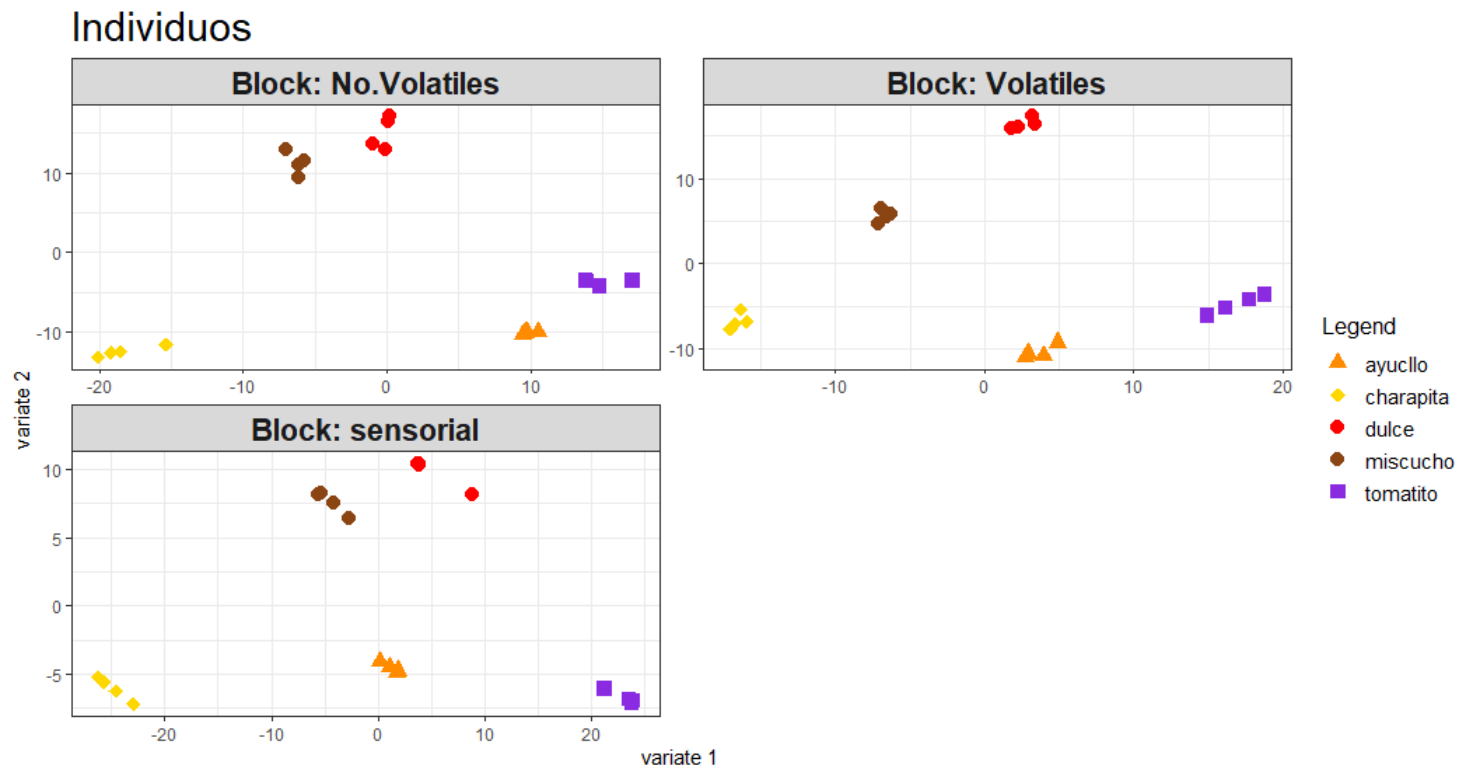


Figura 131. Gráfico de individuos (muestras de ajés) del modelo S-PLSDA de las cinco accesiones de ajés en los tres bloques de datos. Se observa la discriminación de las cinco accesiones de ajés con el mismo patrón de diferenciación en cada bloque.

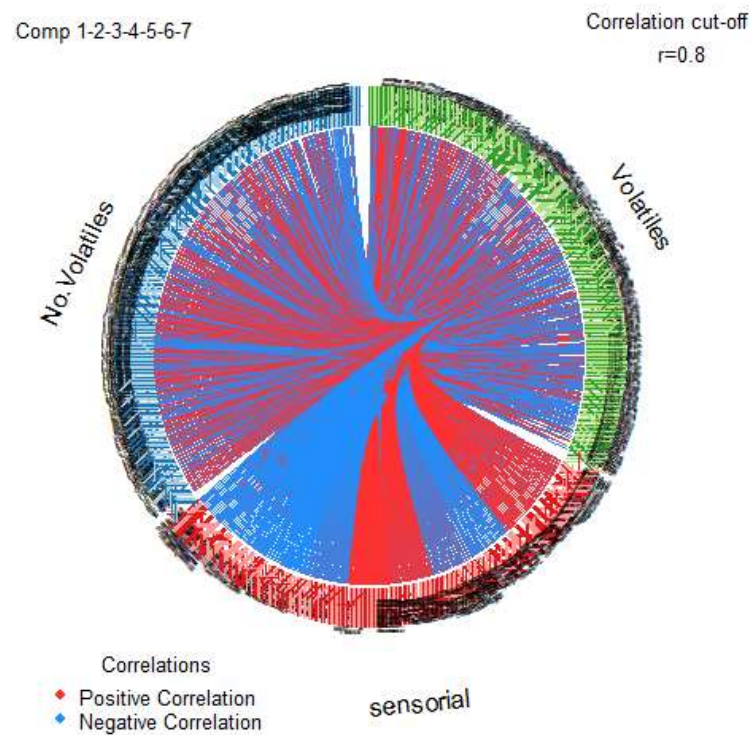


Figura 132. Diagrama Circosplot realizado para la vinculación de la data química (áreas relativas normalizadas GCMS y LCMS) con los scores de los descriptores del sabor y aroma del panel de cata de las cinco accesiones de ajíes nativos. El gráfico muestra las correlaciones positivas y negativas entre los tres bloques de datos. El análisis incluyó un valor de corte cutt-off de $r= 0.8$, para la visualización de los coeficientes de correlaciones por encima de este umbral y la eliminación de las correlaciones más débiles.

Tabla 47. Vinculaciones de los metabolitos volátiles (GCMS) con los descriptores del aroma para la accesión Charapita

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs*	Correlaciones			
				orégano	cítrico	frutal	maracuyá
Fresco	12-Methyl-oxa-cyclododecan-2-one	lactona	**NA	-0.883	-0.969	-0.831	-0.893
Fresco	1-Hexene, 3,5-dimethyl-	alifatico	NA	-0.854	-0.791	-0.959	-0.756
Fresco	1-Penten-3-one, 2-methyl-	cetona	menta	-0.797	-0.858	-0.962	-0.873
Liofilizado	2,4-Dimethyldodecane	alifatico	NA	-0.835	-0.971	-0.823	-0.850
Fresco	4-Methylpentyl 2-methylcrotonate	ester	NA	0.956	0.926	0.914	0.774
Fresco	2-Hexen-1-ol (trans)	alcohol	Verde, afrutado, hoja,	-0.815	-0.876	-0.960	-0.865
Fresco	2-Penten-1-ol, (Z)-	alcohol	Verde, afrutado	-0.962	-0.878	-0.908	-0.649
Liofilizado	2-Undecanone	cetona	Naranja, floral, afrutado, verde	-0.952	-0.961	-0.814	-0.707
Fresco	4-Methylpentyl 8-methylnonanoate	ester	Naranja, floral, afrutado, verde	0.901	0.877	0.751	0.718
Fresco	6-Methyl-4-heptenyl pentanoate	ester	Cálido, rosa, hierbas, miel	0.779	0.948	0.810	0.921
Fresco	6-Octenoic acid, 3,7-dimethyl-, methyl ester	ester	Manzana, afrutado, geranio	0.887	0.958	0.855	0.902
Liofilizado	Benzoic acid, ethyl ester	ester	Uva, cereza, afrutado, fresa	-0.833	-0.986	-0.762	-0.827
Liofilizado	Benzyl isobutyrate	ester	Jazmín, afrutado, tropical	0.635	0.782	0.742	0.709
Liofilizado	Butylated Hydroxytoluene	alcohol	Alcanfor, fenólico, suave	-0.884	-0.892	-0.924	-0.707
Fresco	cis-11-Hexadecenal	aldehido	NA	0.942	0.911	0.907	0.775
Liofilizado	Citral	monoterpeno	Limón, cítricos, menta	-0.873	-0.967	-0.794	-0.778
Fresco	Citronellyl propionate	ester	Neroli, ceroso, rosa, verde, afrutado	0.894	0.993	0.812	0.841
Fresco	citronellyl valerate	sesquiterpeno	Cálido, rosa, hierbas, miel	0.943	0.984	0.865	0.798
Fresco	Cubenene	alifatico	Mango, afrutado, picante	0.625	0.821	0.832	0.945
Liofilizado	Cyclohexene, 4-methyl-1-(1-methylethenyl)-	monoterpeno	NA	-0.896	-0.971	-0.845	-0.872

Tabla 48. Continuación tabla 47...

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs*	Correlaciones			
				orégano	citrico	frutal	maracuyá
Fresco	Cyclopentadecanone, 2-hydroxy-	sesquiterpeno	NA	-0.909	-0.829	-0.980	-0.725
Liofilizado	Cyclopentasiloxane, decamethyl-	N/A	NA	-0.941	-0.967	-0.818	-0.738
Fresco	delta-Decalactone	lactona	Melocotón, coco, albaricoque, afrutado	-0.804	-0.808	-0.960	-0.760
Liofilizado	Dodecane	alifatico	Alcano	-0.832	-0.973	-0.739	-0.835
Fresco	E-14-Hexadecenal	aldehido	NA	0.743	0.901	0.879	0.908
Liofilizado	Ethanone, 1-(3-methylphenyl)-	cetona	NA	-0.806	-0.805	-0.721	-0.848
Liofilizado	Ethyl butyrate	ester	Manzana, piña, plátano, fruta, afrutado	-0.855	-0.961	-0.726	-0.827
Fresco	Heptadecanal	aldehido	NA	0.805	0.938	0.803	0.950
Fresco	Heptyl (E)-2-methylbut-2-enoate	ester	NA	0.906	0.992	0.812	0.815
Fresco	Hexadecane, 2,6,11,15-tetramethyl-	diterpeno	NA	0.911	0.960	0.740	0.765
Liofilizado	Isopropyl butyrate	ester	Piña, madura, frutal	-0.855	-0.890	-0.760	-0.785
Fresco	Ledene oxide-(II)	eter	NA	0.934	0.920	0.875	0.795
Fresco	n-Butyl ether	eter	NA	-0.888	-0.840	-0.983	-0.772
Liofilizado	o-Isopropenyltoluene	monoterpeno	NA	-0.866	-0.964	-0.821	-0.895
Fresco	Oxacyclotetradecan-2-one, 14-methyl-	lactona	NA	-0.868	-0.823	-0.988	-0.754
Fresco	Propanoic acid, 2-methyl-, heptyl ester	ester	Manzana, verde, afrutada, floral, albaricoque, cereza	0.885	0.965	0.844	0.787
Fresco	trans, trans-2,4-Hexadienal	alifatico	cítricos, florales, dulces, picantes, verdes	-0.859	-0.774	-0.894	-0.720
Liofilizado	Undecane, 2,6-dimethyl-	alifatico	NA	-0.862	-0.979	-0.857	-0.875
Liofilizado	Undecane, 4,8-dimethyl-	alifatico	NA	-0.820	-0.971	-0.821	-0.874

Tabla 49. Vinculaciones de los metabolitos No volátiles (LCMS) con los descriptores del Sabor para la accesión Tomatito rojo

Nombre del metabolito	Clase química	Perfil de Sabor ^e , fuente: Publics-DBs*	Correlaciones		
			dulce	pimiento	tomate
2,4',5,6,7,9-Hexahydroxy-8-methylhomoisoflavanone	Flavonoide	NA	-0.843	-0.851	-0.886
2',5,5',7-Tetrahydroxy-6,8-dimethylflavanone; 2'-O-β-D-Glucopyranoside	Flavonoide	NA	0.909	0.919	0.954
3(15),6-Humuladiene-2,9-diol; 9-O-(4-Hydroxy-3-nitro-E-cinnamoyl)	Terpenoide	NA	0.813	0.818	0.898
3,14,16-Trihydroxycard-20(22)-enolide; 16-Ac, 3-O-[β-D-glucopyranosyl-(1→4)-2,6-dideoxy-3-O-methyl-α-L-ribo-hexopyranoside]	Esteroides	NA	-0.979	-0.971	-0.895
3,4,8,10-Tetrahydroxy-1,11(13)-guaiadien-12,6-olide; 8-Ac	Terpenoide	NA	0.956	0.957	0.964
3,6-Dihydroxypregn-9(11)-en-20-one; 3-O-Sulfate	Esteroides	NA	0.845	0.834	0.843
3-Carboxy-1-methylpyridinium betaine	Alcaloide	Hierbas, condimentos y especias	0.909	0.909	0.917
4',6,7-Trihydroxy-8-prenylisoflavone; 4',6-Di-Me ether	Flavonoide	NA	-0.987	-0.969	-0.892
4,6-Decadiyn-1-ol; 4-Hydroxy-3-methoxy-E-cinnamoyl	Alifático	NA	0.948	0.943	0.879
4-Hydroxy-3-methyl-2-butenic acid; Nitrile, O-[2E-butenoyl-(→4)-β-D-glucopyranoside] (Prinsepicyanoside B)	Carbohidrato	NA	-0.951	-0.909	-0.802
4-Hydroxybenzoylcholine	Alcaloide	se encuentra en hierbas y especias, mostaza blanca	0.946	0.930	0.869
4-Hydroxybenzyl alcohol; α-O-(3,4-Dihydroxybenzoyl), 4-O-[4-hydroxy-3,5-dimethoxy-E-cinnamoyl-(→6)-β-D-glucopyranoside]	Aromático	NA	0.837	0.860	0.871
5-O-β-D-Glucopyranuronosyl-L-arabinose; Pyranose-form (Disaccharides)	Carbohidrato	NA	-0.812	-0.821	-0.854
6,8-Di-C-β-D-glucopyranosyl-3',4',7-trihydroxyflavone	Flavonoide	NA	0.958	0.947	0.883
Adenosine; 5'-Ac	Carbohidrato	NA	-0.877	-0.866	-0.862
Aegeline; (+)-form	Alcaloide	NA	-0.940	-0.919	-0.870
Allysine	Aminoácido	NA	-0.931	-0.902	-0.901
Vicenin-2	Flavonoide	En naranjas dulces, albahacas dulces y limones	0.969	0.970	0.937

Tabla 50. Continuación tabla 49...

Nombre del metabolito	Clase química	Perfil de Sabor ^c , fuente: Publics-DBs*	Correlaciones		
			dulce	pimiento	tomate
Cytidine; 4N-Ac	carbohidrato	NA	-0.857	-0.889	-0.964
Edgeworthin; 6-Me ether, 7-Ac	flavonoide	Ginseng y otras bebidas.	-0.824	-0.833	-0.868
Mescaline; N-Ac	Alcaloide	NA	0.894	0.900	0.882
Pentose + Proline	Aminoácido	NA	-0.846	-0.840	-0.829
p-Menthan-3-ol; 2-Aminobenzoyl	terpenoide	NA	0.841	0.855	0.931
Quercetin 3-glycosides; 3-O-[4-Hydroxy-E-cinnamoyl-(→6)-β-D-glucopyranoside]/Rutin	flavonoide	NA	-0.802	-0.843	-0.944
Quercetin 3-glycosides; 3-O-[α-L-Arabinopyranosyl-(1→6)-[4-hydroxy-E-cinnamoyl-(→2)]-β-D-galactopyranoside]	flavonoide	NA	-0.954	-0.957	-0.947
Quercetin 7-glycosides; 7-O-[β-D-Xylopyranosyl-(1→2)-[α-L-rhamnopyranosyl-(1→6)]-β-D-glucopyranoside]	flavonoide	NA	-0.947	-0.953	-0.945
Tyrosine; O-β-D-Glucopyranoside	Aminoácido	NA	-0.806	-0.816	-0.851
Capsaicin; Octanoyl analogue	Alcaloide	Picante, hierbas, condimentos y especias.	-0.844	-0.859	-0.948
Coumaroyl Hexoside	Aromático	NA	-0.803	-0.814	-0.858
Capsaicin; Nonanoyl analogue	Alcaloide	Picante, hierbas, condimentos y especias.	-0.807	-0.831	-0.889

Tabla 51. Vinculaciones de los metabolitos No volátiles (LCMS) con los descriptores del Sabor para la accesión Dulce rojo

Nombre del metabolito	Clase química	Perfil de Sabor ^e , fuente: Publics-DBs [*]	Correlaciones	
			dulce	pimiento
1,5-Bis(3,4-dihydroxyphenyl)-4,5-dihydroxy-1-pentanone; 7'-Me ether, 8'-O-β-D-glucopyranoside	Aromático	NA	0.765	0.757
2,3-Dimethyl-1,4-benzoquinone	Benzoquinona	NA	-0.775	-0.794
2',4',5',7,8-Hexahydroxyisoflavone; 2',4',5',8-Tetra-Me ether	Flavonoide	NA	0.773	0.763
2,6-Nonadienoic acid; Ethylamide	Alcaloide	Sabores y olores a carne	0.771	0.787
3-(2,3-Dihydroxyphenyl)-2-propenoic acid; 3-Me ether	Aromático	Cereales y productos de panadería.	-0.729	-0.754
3,10-Dihydroxy-4,15,17-sphenolobatrien-12,13-olid-19-oic acid; 10-Ac, Me ester	Terpenoide	NA	-0.765	-0.702
3,3',4,4'-Tetrahydroxy-8,8'-lignan-9,7':9',7'-diolide; 3,3'-Di-Me ether	Lignano	Cereales y productos de panadería.	-0.784	-0.791
3,5,14-Trihydroxycard-20(22)-enolide; 3-O-[β-D-Glucopyranosyl-(1→4)-2,3-di-O-methyl-β-D-fucopyranoside]	Esteroides	NA	-0.794	-0.761
3,5,7,8,9,15-Hexahydroxy-6(17),11-jatrophadien-14-one; 8-O-(2-Methylbutanoyl), 5,7-bis-O-(methylpropanoyl)	Terpenoide	NA	0.763	0.771
3,5,7-Trihydroxyflavanone; 3-O-(3,4-Dihydroxy-E-cinnamoyl)/7-hydroxy-2-(4-hydroxyphenyl)-5-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy-2,3-dihydrochromen-4-one	Flavonoide	en, frutas y prunus (cereza, ciruela).	-0.772	-0.748
3,8,11,12,14-Pentahydroxypregnan-20-one; 12-Benzoyl, 8,11-ditigloyl, 3-O-[6-deoxy-3-O-methyl-β-D-allopyranosyl-(1→4)-6-deoxy-β-D-glucopyranoside]	Esteroides	NA	0.746	0.760
3-O-α-L-Arabinofuranosyl-D-xylose; Me glycoside	Alcaloide	NA	0.733	0.752
3-Phenyl-2-propen-1-ol; O-[α-L-Rhamnopyranosyl-(1→6)-β-D-glucopyranoside]	Aromático	NA	0.778	0.780
4'-Hydroxyacetophenone	Aromático	Sabores y olores dulces, café	0.777	0.794
6'',7''-dihydro-5',5'''-dicapsaicin	Alcaloide	no picante	0.766	0.776
8,14-Epoxy-3,4-dihydroxycardanolide; 20,22-Didehydro, 4-Ac, 3-O-[β-D-glucopyranosyl-(1→4)-α-D-oleandropyranoside]	Esteroides	NA	-0.754	-0.731

Tabla 52. Continuación tabla 51...

Nombre del metabolito	Clase química	Perfil de Sabor ^e , fuente: Publics-DBs*	Correlaciones	
			dulce	pimiento
Capsaicin; 7-Methyl-5E-octenoyl analogue	Alcaloide	Picante, hierbas, condimentos y especias.	0.796	0.759
Dobutamine	Alcaloide	NA	0.772	0.769
Ergosta-5,24-diene-1,3,22,26-tetrol; 26-O-β-D-Glucopyranoside	esteroide	NA	0.754	0.770
Ethyl-1,4-benzoquinone	Benzoquinona	NA	0.787	0.755
Homotyramine; Me ether, N-(3-hydroxy-4-methoxy-E-cinnamoyl)	alcaloide	NA	0.735	0.705
Jujuphenoside	Aromático	NA	0.726	0.772
Luteolin-4'-O-glucoside/Quercetin-3-O-rhamnoside	flavonoide	NA	-0.703	-0.731
Macaflavone I; 3'-Deoxy, 4'-Me ether	flavonoide	NA	0.721	0.702
Methionine	Aminoácido	NA	-0.783	-0.781
N6-(5-Amino-5-carboxypentyl)-5-hydroxylysine	aminoácido	NA	0.785	0.759
N-benzoyl-4-hydroxyanthranilic acid	Alcaloide	NA	0.726	0.711
Phosphocholine	Alcaloide	toronja, lima, repollo negro y cebada	-0.785	-0.796
p-Mentha-3,8-dien-1-ol; (β)-form	terpenoide	Hierbas, condimentos y especias	0.763	0.780
Spirostane-1,2,3-triol; 1-O-β-D-Glucopyranoside	esteroide	NA	0.715	0.725
Torosafavone D; 4'-O-De-Me, 3'-deoxy, Me ester/Chlorogenic acid	flavonoide	NA	-0.753	-0.790
8-Hydroxy-9,10-dinor-4-megastigmen-3-one; O-β-D-Glucopyranoside	Terpenoide	NA	-0.728	-0.720
Capsaicin	Alcaloide	Picante, hierbas, condimentos y especias.	0.749	0.765

Tabla 53. Vinculaciones de los metabolitos Volátiles (GCMS) con los descriptores del Aroma para la accesión Dulce rojo

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs [*]	Hierba
Fresco	1,16-Hexadecanediol	alcohol	NA	-0.837
Fresco	1,1'-Bicyclopentyl, 2-hexadecyl-	alifatico	NA	0.738
Fresco	1-Butanol, 3-methyl-, acetate	ester	amargo, solvente, plátano, afrutado	-0.750
Fresco	1H-Cycloprop[e]azulene, 1a,2,3,4,4a,5,6,7b-octahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.alpha.,4a.beta.,7b.alpha.)]-	sesquiterpeno	Madera, bálsamo	-0.869
Fresco	1-Pentanol, 4-methyl-	alcohol	sudor	-0.820
Fresco	2(3H)-Furanone, 5-ethylidihydro- ó gamma-Caprolactone	lactona	coco, tabaco, cumarina, herbal, dulce	0.730
Fresco	2-Decanol	alcohol	NA	0.741
Liofilizado	2H-Pyran-2-one, 5,6-dihydro-6-pentyl-	lactona	Coco, carne, cremoso lactónico, lechoso, ceroso	0.716
Fresco	2-Pentenal, (E)-	aldehido	manzana, naranja, verde	-0.866
Fresco	2-Tetradecanone	cetona	NA	0.758
Fresco	2-Tridecanone	cetona	herbal, graso, terroso, lácteo, nuez	0.765
Fresco	4-Aminobutanoic acid	aminoacido	NA	0.821
Fresco	5H-1-Pyridine	piridina	NA	0.831
Fresco	Benzene, 1,3-dimethyl-5-(1-methylethyl)-	aromatico	NA	-0.827
Fresco	beta-cubebene	sesquiterpeno	afrutado, cítrico, fruta	-0.776
Liofilizado	Butanoic acid, 3-methyl-, 3,7-dimethyl-6-octenyl ester	ester	manzanilla, manzana, grosella negra, rosa, hierbas	0.748
Fresco	Dodecanoic acid, 2-hexen-1-yl ester	ester	NA	-0.739
Liofilizado	Ethanol, 2-phenoxy-	eter	NA	0.743

Tabla 54. Continuación tabla 53...

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs*	Correlaciones
				hierba
Fresco	Isoamyl alcohol	alcohol	NA	-0.734
Liofilizado	Methanesulfonylacetic acid	acido sulfonico	NA	0.749
Fresco	Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)-(.)-	sesquiterpenos	Verde, leñoso	-0.763
Fresco	Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	sesquiterpenos	Verde, leñoso	-0.774
Fresco	Oxacycloheptadecan-2-one	lactona	NA	-0.785
Fresco	Pentanoic acid, 4-methyl-, phenylmethyl ester	ester	NA	0.902
Fresco	Phthalic acid, di(2-propylpentyl) ester	ester	NA	-0.841
Liofilizado	Ethyl laurate; ethyl decanoate	ester	Dulce, floral ceroso, jabonoso	0.771

Tabla 55. Vinculaciones de los metabolitos Volátiles (GCMS) con los descriptores del Aroma para la accesión Miscucho

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a,b,c,d} , fuente: Publics-DBs*	Correlaciones		
				Manzana	Frutal	Maracuyá
Liofilizado	1-Octanol, 2-butyl-	alcohol	NA	-0.626	-0.696	-0.816
Liofilizado	2,4-Di-tert-butylphenol	fenol	NA	-0.667	-0.652	-0.689
Fresco	Butanoic acid, 3-methyl-, 2-phenylethyl ester	ester	rosa, durazno, albaricoque, floral, afrutado	0.706	0.626	0.762
Liofilizado	Cyclohexane, 1-methylene-4-(1-methylethenyl)-	monoterpenos	NA	-0.791	-0.678	-0.835
Fresco	Furan, 2-pentyl-	furano	Afrutado	-0.663	-0.968	-0.619
Liofilizado	Heptanoic acid, methyl ester	ester	Fruta, baya	-0.619	-0.676	-0.965
Liofilizado	Heptanoic acid, propyl ester	ester	Notas frutales, de manzana, pera, piña, uva y fresa.	-0.641	-0.644	-0.761
Liofilizado	Hexanoic acid, 4-hexen-1-yl ester	ester	NA	-0.652	-0.650	-0.773
Fresco	nonyl isovalerate	ester	manzana, afrutado, cítrico	0.613	0.946	0.697
Fresco	Octanoic acid, hexyl ester	ester	manzana, afrutado, verde	0.639	0.686	0.814

Tabla 56. Vinculaciones de los metabolitos Volátiles (GCMS) con los descriptores del Aroma para la accesión Ayucllo

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs*	Correlaciones	
				Hierba	Manzana
Fresco	3,5-Dimethyl-2-octanone	cetona	NA	-0.838	-0.859
Fresco	3,7-Dimethyl-1,3,6-octatriene; ocimene	monoterpeno	cítricos, herbales, dulces, leñosos, tropicales, terpenos, hierbas, verde	0.838	0.852
Fresco	3-Hexenal	aldehido	frondoso, manzana, tallo, tomate, melón, afrutado, natural, verde	-0.766	-0.776
Fresco	4-Methylpentyl 8-methylnon-6-enoate	ester	NA	0.880	0.957
Fresco	Acetic acid, 2-ethylhexyl ester	ester	NA	-0.823	-0.838
Fresco	Benzene, 1-methyl-4-(1-methylpropyl)-	aromatico	NA	0.767	0.728
Fresco	Benzyl butyl phthalate	ester	NA	-0.895	-0.969
Fresco	beta-Humulene	sesquiterpeno	NA	-0.901	-0.858
Fresco	Bicyclosquiphellandrene	sesquiterpeno	NA	-0.894	-0.904
Fresco	Butanoic acid,4-hexen-1-yl ester	ester	NA	-0.839	-0.772
Fresco	Chlorpyrifos	NA	NA	-0.864	-0.757
Fresco	cis-Calamenene	sesquiterpeno	NA	-0.834	-0.815
Fresco	dioctyl adipate	ester	graso, leve	0.918	0.933
Fresco	E-11(13-Methyl) tetradecen-1-ol acetate	ester	NA	0.904	0.911
Liofilizado	Fenchyl alcohol	monoterpeno	limón, lima, pimienta, uva, mohoso, alcanforado, frambuesa, alcanfor	-0.745	-0.784
Fresco	Germacrene D	sesquiterpeno	madera, especias	0.715	0.709
Fresco	Guaiyl acetate	alifatico	NA	-0.841	-0.768
Fresco	Hexadecanal	aldehido	cartulina	0.851	0.963

Tabla 57. Continuación tabla 56...

Muestra de ají	Nombre del metabolito	Clase química	Perfil de Aroma ^{a, b, c, d} , fuente: Publics-DBs*	Correlaciones	
				Hierba	Manzana
Fresco	Hexanoic acid, 4-pentenyl ester	ester	NA	0.847	0.773
Liofilizado	Methyl 8-methyl-nonanoate	ester	NA	0.831	0.742
Fresco	N,N,O-Triacetylhydroxylamine	acetamida	NA	0.863	0.731
Fresco	Pentadecanal-	aldehido	Ceroso, fresco	0.798	0.932
Fresco	Phenylethyl methacrylate	ester	NA	-0.804	-0.736
Fresco	trans-Geranic acid methyl ester	ester	Ceroso, fruta, afrutado, flor, verde	0.827	0.820
Fresco	Tridecane	alifatico	Alcano	0.809	0.862
Fresco	Tridecyl acetate	ester	NA	0.933	0.916
Fresco	Valencene	sesquiterpeno	NA	0.890	0.921
Fresco	Hexadecane	alifatico	alcano	0.719	0.903

Tabla 58. Vinculaciones de los metabolitos No Volátiles (GCMS) con los descriptores del Sabor para la accesión Ayuclo

Nombre del metabolito	Clase química	Perfil de Sabor ⁶ , fuente: Publics-DBs ⁸	Correlaciones
			Acidez
14,17,20,27-Tetrahydroxy-1-oxowitha-3,5,24-trienolide; (14a,17 β ,20S,22R)-form	esteroide	NA	0.969
15,20-Dihydroxypregn-4-en-3-one; 20-O-[β -D-Glucopyranosyl-(1 \rightarrow 4)-6-O-acetyl- β -D-glucopyranoside]	esteroide	NA	0.906
2-(4-Methoxyphenyl) ethylamine; N-E-Cinnamoyl	Alcaloide	NA	-0.817
2,3-Dihydro-5,7-dihydroxy-2-methyl-4H-1-benzopyran-4-one	benzopirano	NA	0.857
3-(3,4-Dihydroxyphenyl)-2-propen-1-ol; 3'-Me ether, 1-cinnamoyl	Aromático	NA	-0.831
3-(3,4-Dihydroxyphenyl)-2-propen-1-ol; 3'-Me ether, 4'-sulfate	Lignano	NA	0.893
3,16-Dihydroxyergosta-8,16-diene-15,23-dione; 24,28-Didehydro, 3-O-[β -D-xylopyranosyl-(1 \rightarrow 3)- β -D-glucuronopyranoside]	esteroide	NA	0.835
3,16-Dihydroxyergosta-8,16-diene-15,23-dione; 3-O-[β -D-Xylopyranosyl-(1 \rightarrow 3)- β -D-glucuronopyranoside]	esteroide	NA	0.819
3,4',5,7,8-Pentahydroxy-6-methylflavone; 3-Me ether, 4',8-di-O- β -D-glucopyranoside	flavonoide	NA	-0.838
3,4-Dihydroxybenzoic acid; 3-Me ether, 4-O-[3,4-dihydroxy-E-cinnamoyl-(\rightarrow 6)- β -D-glucopyranoside]	Aromático	NA	-0.898
3,5,10,14-Tetrahydroxy-19-norcard-20(22)-enolide; 3-O-(4,6-Dideoxy-2-O-methyl- β -D-ribohexopyranoside)	esteroide	NA	0.905
3,5,14-Trihydroxy-19-oxocard-20(22)-enolide; 3-O-[α -L-Rhamnopyranosyl-(1 \rightarrow 4)-3-O-acetyl-2,6-dideoxy- β -D-ribohexopyranoside]	esteroide	NA	0.922
3,5,14-Trihydroxycard-20(22)-enolide; 3-O-[β -D-Glucopyranosyl-(1 \rightarrow 4)-2,3-di-O-methyl- β -D-fucopyranoside]	esteroide	de la corteza de la fruta de <i>Streblus asper</i>	0.819
3,8,9-Trihydroxy-2-prenylcoumestan; 8,9-Methylene ether/sucrose	flavonoide	NA	-0.885
5,20-Epoxy-11-taxene-1,2,4,7,9,10,13-heptol; 2-Hexanoyl, 4,7,9,10,13-penta-Ac	terpenoide	NA	0.841
5,7-Dihydroxy-8-(4-hydroxyphenyl)-2,2-dimethyl-2H,6H-benzo[1,2-b:5,4-b']dipyrano-6-one; 3'-Hydroxy, 3-Me ether/ Galactinol	flavonoide	NA	-0.876
7-Megastigmene-3,6,9-triol; 9-O-[4-Hydroxy-3,5-dimethoxybenzoyl-(?5)-?-D-apiofuranosyl-(1?6)-?-D-glucopyranoside]	terpenoide	NA	-0.826

Tabla 59. Continuación tabla 58...

Nombre del metabolito	Clase química	Perfil de Sabor*, fuente: Publics-DBs*	Correlaciones
			Acidez
8,9-Dihydroxy-14-oxo-1(10),4,11(13)-germacatrien-12,6-olide; 9-O-(2-Hydroxy-2-methylpropanoyl), 8-O-(2-methylpropanoyl)	terpenoide	NA	0.863
8,9-Dihydroxy-3,10(14),11(13)-guaiatrien-12,6-olide; 8-O-(2-acetoxymethyl-2E-butenoyl), 9-Ac	terpenoide	NA	0.804
9,11,13-Octadecatrienoic acid; 2,3-Dihydroxypropyl ester/Monolinolenin (9c-12c-15c)	Alifático	Lipido, encontrado en frutas	0.896
9,12,13,20-Tetrahydroxy-1,6-tigliadien-3-one; 12-Hexanoyl, 13-Ac	terpenoide	NA	0.907
9-Hydroxy-1(10),4,11(13)-germacatrien-12,6-olide; O-β-D-Glucopyranoside	terpenoide	NA	-0.806
Adenosine	carbohidrato	NA	0.959
Bonaspectin D; 4"-O-β-D-Glucopyranoside	lignano	NA	0.896
Capsicum annum Fluorescent chlorophyll catabolite	polipirrol	colorante, en hierbas, condimento y especias	0.901
Glutamylleucine	Aminoácido	encontrado en ajos	0.905
Lysine; 2-N-Ac	Aminoácido	NA	0.902
N-?-Glutamylleucine	Aminoácido	NA	0.916
Salicin; 1"-O-(4-Hydroxy-3-methoxy-E-cinnamoyl)	Aromático	NA	-0.839
Streptospirdienoic acid A; Me ester	poliketido	NA	0.815
Withaperuvin H	esteroide	encontrado en frutas	0.822

**NA = no disponible en las bases de datos públicas

*Bases de datos públicas (Publics-DBs) de aromas y sabores:

a) <http://www.thegoodscentscompany.com>, b) <https://cosylab.iiitd.edu.in/flavordb/search>, c) <https://www.flavornet.org/index.html>, d) <http://www.perflavory.com/allflavor.html>, e) <http://foodb.ca/>

8.DISCUSIÓN DE RESULTADOS

8.1.Análisis sensorial de las cinco accesiones de ajíes nativos

Para evaluar la confiabilidad de la data sensorial se utilizó una herramienta estadística denominada análisis de Panel del software XLSTAT de Excel. Esta herramienta comprueba mediante un análisis ANOVA si el panel puede resaltar (y en qué medida) las diferencias entre los descriptores y los panelistas. El análisis de varianza ANOVA es un análisis común en el análisis sensorial con aplicaciones en diferentes alimentos tales como café (130) y jugos de maracuyá (131). Según los resultados del ANOVA, existe una diferencia significativa tanto entre los panelistas como entre las muestras. Un efecto se considera significativo si el valor p resultante es pequeño (generalmente el valor de 0.05 se toma como umbral) (132). En general, el Panel mostró una reproducibilidad moderada de las percepciones del sabor y aroma de los ajíes. Del análisis de ANOVA, solo el descriptor “hierba” no fue bien modelado. Para rechazar la hipótesis nula de que las medias grupales son iguales es necesario un alto valor de F (F -test) (133). El descriptor “hierba” presenta el valor F más bajo ($F = 1.385$). También, el valor de $p = 0.157 > 0.05$, (siendo $p = 0.05$ el nivel de significancia) calculado indicaría que se acepta la hipótesis nula para este descriptor. Es decir, hay una diferencia significativa entre los panelistas, y entre las muestras también hay una diferencia significativa, excepto por el descriptor hierba. El descriptor más fuertemente modelado es el descriptor tomate ($F = 12.146$, $p < 0.001$) y los más débilmente modelados son los descriptores manzana ($F = 1.744$,

$p = 0.045$) y orégano ($F = 1.806$, $p = 0.036$). También se observó que el poder discriminante de los panelistas varía en el rango de 33-89%. La falta de discriminación del descriptor “hierba” pudo deberse al bajo poder discriminante por parte del panelista 7 (33%). De acuerdo con los resultados, el descriptor hierba está presente y aunque no fue bien modelado, no ha sido descartado como descriptor. También se pudo observar que no existe una homogeneidad de las percepciones por parte de los panelistas ya que cada panelista presenta diferente poder discriminante. Esta heterogeneidad de la percepción de los sabores y aromas por los catadores puede ser debido a las propias subjetividades de los panelistas (134), o por la presencia de la Capsaicina que le da una sensación de pungencia a los ajíes, lo cual podría enmascarar algunas notas del aroma. Otro aspecto podría ser la concentración variable del metabolito que produjo la sensación en los diferentes ajíes, debido a diferente grado de madurez, o diferentes grados de volatilidad de los aromas por cambios en la temperatura ambiental.

Por otro lado, existen opiniones muy divergentes sobre el número de panelistas necesarios para el análisis descriptivo sensorial (QDA), pero en general se recomienda trabajar entre 8 a 10 panelistas (135). En este estudio se trabajó con 9 panelistas y fue suficiente para el trabajo de análisis sensorial. El tamaño de la data sensorial fue lo suficientemente adecuada para un tamaño muestral con distribución normal (27 datos por cada descriptor). Asimismo, el panel paso por una etapa de entrenamiento previo de un año por su participación del proyecto FINCYT por lo que fue posible obtener resultados reproducibles. Una dificultad del estudio fue la

disponibilidad de ajíes frescos ya que estos sufren una rápida maduración. Los ajíes solo se cosechan en la estación de verano (una vez al año), por lo que las pruebas se realizan solo en esta época. Son cosechas únicas de ajíes nativos que han pasado por un proceso de adaptación y un manejo agrícola especial. Los ajíes una vez cosechados fueron almacenados adecuadamente y evaluados inmediatamente para evitar perder la información sensorial debido a su rápida maduración y degradación por las altas temperaturas de la estación de verano (136).

Se han identificado los siguientes tres aspectos del análisis sensorial: 1) las puntuaciones relativas (scores) de los descriptores para una accesión dada deben describir su perfil sensorial, 2) el QDA permite obtener la densidad de probabilidad para los descriptores del sabor y aroma, 3) del QDA se puede desarrollar un modelo sensorial multivariado para las cinco accesiones de ajíes. En todos los casos de análisis de data para cada descriptor de una accesión dada, la data sensorial no mostró una distribución normal cuando se ensayaron con los test de normalidad. En ese sentido fue necesario realizar un tratamiento adecuado de la data sensorial. Para ello, se utilizó la definición del teorema del límite central, en el que la distribución muestral de la media muestral, \bar{X} , para un tamaño de muestra grande, es casi normal, sea cual sea la forma de la distribución de la población subyacente (con una media igual a μ , que es la media de la población, y error estándar al cuadrado igual a $(1/n)$ (σ^2), que es la varianza de la población dividida entre n) (137). Además, a medida que el tamaño de la muestra se incrementa, la media muestral se acercará más a la media poblacional.

Tomando en cuenta este teorema, se consideró que cada uno de los scores de los descriptores del sabor y aroma para una accesión dada con $N= 27$ datos, provienen de una media muestral, y que en conjunto tienen una distribución normal. Es decir, los 27 datos son en realidad 27 medias muestrales. Para justificar esta premisa se debe tomar en cuenta que el aroma de los ajíes es una mezcla de metabolitos y presentan los atributos sensoriales volátiles que los panelistas perciben y lo expresan mediante un score de aroma. De igual manera, el sabor comprende metabolitos volátiles y no volátiles que como un todo es una mezcla compleja. En ese sentido, cada score de puntuación representa la suma de sensaciones de un conjunto de metabolitos y no un metabolito individual en específico. La distribución de ese conjunto de sensaciones es desconocida. Por consiguiente, de acuerdo al teorema del límite central, la media muestral de los scores de un descriptor dado debe ser aproximadamente igual a su media poblacional para ese descriptor y su distribución debe ser aproximadamente normal. Los únicos parámetros que se mantienen aproximadamente constantes en los descriptores son los parámetros de la media y desviación estándar. Estos parámetros se utilizaron para obtener la data sensorial artificial y construir un modelo sensorial multivariado.

Las evaluaciones preliminares de análisis multivariados de PCA y OPLS-DA de la data sensorial original arrojaron resultados inapropiados para concluir con una tendencia clara de clasificación categórica por especies de ajíes, ya que los

parámetros de calidad tales como R^2Y , Q^2Y (del modelo OPLS-DA) demostraron una pobre calidad de ajuste y de predicción.

La data artificial fue obtenida para $N= 300$ datos, utilizando el software Autograph v.4.0. El software permitió generar una data utilizando los parámetros de la media y la desviación estándar de la data original. La data artificial tuvo un comportamiento normal cuando se analizaron con los test de Normalidad y fue la información base del modelo. Con esta data, se desarrollaron modelos multivariados de PCA y OPLS-DA utilizando el software SIMCA 15. El PCA es un método de reconocimiento de patrones y proporciona una información visual de cuan bien se empaquetan las muestras para formar cluster entre las muestras. Es una técnica multivariada para exploración y análisis de datos que proyecta los datos en un hiperespacio reducido, definido por los componentes principales. Estos últimos son combinaciones lineales de las variables originales con el primer componente principal con la mayor varianza, y el segundo componente principal con la segunda mayor varianza, y así sucesivamente (138). Mientras que el OPLS-DA es un método adecuado para la clasificación de las diferentes clases haciendo el uso apropiado del análisis discriminante para fines de predicción (139).

El modelo multivariado de PCA de la data sensorial optimizada mostró una clara tendencia de formación de cluster de muestras en función de la clase especie. Esta tendencia fue confirmada en el modelo OPLS-DA en la discriminación visual de las clases. La creación de un modelo tiene que pasar por un proceso de validación.

Para validar el modelo fue necesario utilizar diferentes pruebas. Entre estas pruebas se encuentra el modelo OPLS-DA de la data sensorial optimizada que arrojó un mejor clusterizado y clasificación por especies que el modelo OPLS-DA original. La fiabilidad de la predicción es importante en la evaluación del rendimiento de un modelo de regresión. Esto se vio reflejado por sus valores de R^2Y y $Q^2Y > 0.7$ que indicaron una buena calidad de ajuste y predicción. Los test de área bajo la curva de los gráficos de ROC mostraron una excelente capacidad de clasificación por especies, y adicionalmente el test de Permutaciones y CV-ANOVA permitieron validar el modelo internamente. Además de los procedimientos clásicos de validación interna, como la validación cruzada, se utilizó la validación mediante un conjunto de prueba externo. La data utilizada para esta validación externa fue la misma data sensorial original ya que fue la única data disponible externa a la data optimizada que no se utilizó para construir el modelo OPLS-DA. Los resultados mostraron confiables valores de probabilidad > 0.5 , lo que confirma la capacidad del modelo de predecir nuevos scores del modelo OPLS-DA con una alta precisión. Una vez validada tanto interna como externamente, la data fue lo suficientemente robusta para la predicción de nuevas muestras. Existen pocos ejemplos de diseño de modelos sensoriales. Se puede citar el modelo sensorial desarrollado por Lindinger *et al.* (140) para el café expreso, para predecir el perfil sensorial a partir de data instrumental de metabolitos volátiles. En este trabajo, en el análisis sensorial se utilizaron 10 panelistas, utilizando el enfoque del QDA. Se evaluaron nueve atributos de aroma y dos sabores en escala no estructurada de 0-10. La capacidad de discriminación fue probada para cada atributo utilizando un ANOVA de dos vías

con el producto como valor fijo y el panelista como factor aleatorio. La relación F (test de Fisher) del efecto del producto sirvió como marcador de la capacidad de discriminación. El desempeño individual de un evaluador en un solo atributo consistió en la capacidad de discriminación individual y la capacidad de ser consensado con los otros evaluadores del panel. La capacidad de discriminación individual fue probada usando un ANOVA de una vía con el producto como factor fijo. Se aplicó la regresión de componentes principales (PCR) para obtener la matriz que transforma los datos instrumentales en datos sensoriales predichos. Los resultados de las predicciones se mostraron en gráficos de radar. Las diferencias encontradas entre este modelo con el modelo sensorial de los ajíes nativos propuesto en este trabajo, se encuentra en el modelo de regresión empleado. Por un lado, el método de PCR está basado en un modelo de análisis de componentes principales, un PCA aplicado a regresión lineal. El PCA es un método no supervisado que no permite evaluar la capacidad predictiva del modelo y solo permite evaluar su capacidad de ajuste. En cambio, el modelo sensorial de ajíes nativos se basa en el modelo supervisado OPLS-DA que es un modelo más potente porque permite evaluar en el modelo su capacidad de ajuste, predicción y análisis discriminante.

En la actualidad no existen reportes sobre estudios similares a los utilizados por este enfoque de modelado de data sensorial. Los estudios en *Capsicum* y otros alimentos se limitan a aplicar la data sensorial tal cual se ha obtenido y no consideran la capacidad predictiva de la misma. En ese sentido, es la primera vez que se reportan estos resultados.

Por otro lado, el desarrollo del perfil sensorial implica el uso de diagramas de radar para la descripción sensorial. Alternativamente se ha utilizado el diagrama de barras para describir el perfil sensorial en pastas de cacao (141), que en este trabajo se ha aplicado para describir el perfil sensorial de los ajíes. El perfil sensorial en gráficos de barras ofrece una descripción más comprensible de los scores del sabor y aroma cuyos resultados fueron obtenidos de la data sensorial optimizada. El análisis del perfil sensorial de los ajíes generó un perfil de puntuaciones relativas de notas de sabores y aromas de los ajíes. Esto podría servir como una guía para la selección objetiva de atributos específicos que podrían tener un impacto en las preferencias de los consumidores. El análisis sensorial descriptivo resultante proporcionó una “imagen” sensorial bidimensional de los ajíes. El test de Tukey permitió saber cuáles son las diferencias significativas observadas por el ANOVA.

La información proporcionada por el gráfico de barras y análisis de varianza ANOVA esta dirigida a identificar niveles de intensidades relativas, mas no permite explicar las relaciones existentes entre las variables. Un análisis multivariado por AFM (Análisis Factorial Múltiple) proporcionó esta información. El AFM permitió explicar cuáles son las variables más importantes en la data sensorial y qué variables estaban correlacionadas mutuamente. Además, el AFM de la data sensorial optimizada y normalizada por el método EigenMS permitió observar las misma separación de clases que en el modelo OPLS-DA de la data sensorial optimizada. Para la interpretación de los resultados del AFM se debe considerar que los descriptores del sabor y aroma dentro del círculo de correlaciones son vectores. Las correlaciones se identifican con el valor del producto punto de dos vectores, es

decir: $\vec{a} \times \vec{b} = a \cdot b \cos \theta$. Si los vectores forman un ángulo agudo ó $\theta = 0^\circ$ entre ellos entonces se dice que los descriptores están fuertemente correlacionados. Si forman 90° entre ellos entonces no están correlacionados. En el círculo de correlaciones se observa que los descriptores dulce, pimiento y tomate están fuertemente correlacionados (los vectores forman un ángulo agudo entre ellos) (129). Para el análisis por AFM, la data sensorial fue analizada por localidad. Solo se analizó por AFM la data sensorial de la localidad de Chincha. Los resultados del AFM de la data sensorial de la localidad de Chiclayo y Huaral son ilustrativas y muestran las mismas correlaciones que el AFM de la data sensorial de la localidad de Chincha. Los resultados del AFM de la data instrumental (análisis proximal, pH y % Brix) de Chincha, en la gráfica de individuos permitió la evaluación de la calidad del modelo de predicción basado en datos instrumentales. Los resultados muestran una excelente calidad en la predicción y sorprendentemente mostró la misma tendencia observada en la gráfica de individuos del AFM de la data sensorial de Chincha. Los resultados del AFM también permitieron inferir sobre la capacidad de discriminación de los descriptores por parte del panel. Según el AFM, el panel modela bien los descriptores de tomate, pimiento y dulce, mientras que los más débilmente modelados son los descriptores de hierba, manzana y orégano.

De la data sensorial optimizada también se construyó un diagrama Heatmap. Los diagramas Heatmap son útiles para visualizar datos multivariados. Tales datos multivariados incluyen data proteómica y metabolómica (142,143). Aplicado a la data sensorial, los diagramas Heatmap han permitido esclarecer aún más el

comportamiento sensorial de los ajíes. Para análisis de datos sensoriales no se reportan este tipo de análisis. Los resultados muestran una clara clasificación por accesiones y especies y permite evaluar los descriptores del sabor y aroma para cada una de las accesiones.

Un análisis en tres dimensiones (3D) de la data sensorial por PLS-DA, permitió evaluar su capacidad de ajuste y predictiva (R^2Y y $Q^2Y > 0.9$). El análisis permitió evaluar los valores de importancia (VIP). Estos valores $VIP > 1$ mostraron los descriptores sensoriales más discriminantes para cada accesión. Esta información es complementaria al AFM, Heatmap que en conjunto ofrecen una visión holística del análisis sensorial.

8.2. Diferencias metabolómicas encontradas en la cosecha orgánica y convencional

La creciente demanda de alimentos orgánicos se debe a las preferencias de los consumidores en productos de calidad nutritivos y de seguridad alimentaria, y la idea de que los alimentos producidos orgánicamente son más saludables y seguros que los alimentos convencionales (144). En la actualidad existe una amplia gama de productos orgánicos y subproductos procesados tales como melocotones, brócoli, tomates, pimiento, peras, ciruelas amarillas y zanahorias (145). Las diferencias fundamentales entre los sistemas de producción orgánicos y convencionales, se centran en el manejo de la fertilidad del suelo que

puede afectar la composición nutritiva de las plantas y a los metabolitos secundarios. La agricultura orgánica es un sistema de manejo de producción ecológico que promueve y mejora la biodiversidad, los ciclos biológicos y la actividad biológica del suelo. Los sistemas orgánicos dependen de la actividad de un ecosistema de suelo diverso para hacer que el nitrógeno y otros nutrientes estén disponibles para las plantas, mientras que los sistemas convencionales utilizan fertilizantes que contienen nitrógeno inorgánico soluble y otros nutrientes, que están más directamente disponibles para las plantas (146). El conocimiento actual de comparación entre estos sistemas se centra en cuatro objetivos de sostenibilidad: productividad, salud ambiental, viabilidad económica y calidad de vida. En los cultivos orgánicos a pesar de la evidencia de menores rendimientos, se reportan evidencias considerables de beneficios ambientales y sociales (147).

Los metabolitos secundarios cumplen funciones importantes para la salud humana. Uno de los mas grandes grupos de metabolitos en plantas se encuentran los compuestos fenólicos tales como los flavonoides y los terpenoides (carotenoides y xantofilas). Estos juegan un rol importante en la prevención del desarrollo de diversas enfermedades tales como el cáncer y la aterosclerosis (148–150). Cumplen funciones en el fortalecimiento de las paredes de los vasos sanguíneos y reducción al riesgo de accidente cerebrovascular (151). También apoyan al sistema inmunológico en la prevención de infecciones por virus y bacterias (152). El segundo grupo de metabolitos secundarios son los alcaloides, aminas, aminoácidos,

glucósidos y glucosinolatos no proteicos. Su impacto en la salud humana es menos conocido que el grupo de sustancias descrito anteriormente (153,154).

Un análisis reciente de 343 investigaciones sobre la composición química de los cultivos orgánicos y convencionales mostró que el contenido de una serie de antioxidantes clave en las materias primas orgánicas y sus productos es hasta un 60% más alto que en plantas cultivadas convencionalmente. En el caso de los carotenoides, este análisis mostró un contenido significativamente mayor en frutas orgánicas que en frutas convencionales; para hortalizas y cereales, las diferencias no fueron estadísticamente significativas. Numerosas investigaciones sobre pimientos y tomates indican que el nivel de luteína y beta-caroteno suele ser más alto en cultivos orgánicos que en los convencionales, mientras que el nivel de licopeno presenta una situación inversa (155) .

Para el caso de las 5 accesiones de ajíes nativos, el análisis metabolómico de las cosechas orgánicas y convencional arrojaron algunas diferencias. Por un lado se observa en el PCA la separación de clases entre ambas cosechas orgánica vs convencional. Sin embargo, no se observa diferencias significativas entre los perfiles LCMS en ambas cosechas ya que el contenido de metabolitos es el mismo pero sus concentraciones son diferentes. La principal diferencia se refleja en el contenido de Capsaicina, siendo este mayor en la cosecha convencional. Del análisis OPLSDA (figuras 81, 82 y 83), se puede observar diferencias en la separación de cluster de las cosechas orgánica y convencional, esto corrobora la

hipótesis propuesta en el PCA. El análisis S-plot mostró los metabolitos marcadores que discriminan ambas cosechas. Las tablas 34, 35 y 36 mostraron los resultados del análisis S-plot para las tres especies. Sin embargo, este análisis no dio resultados concluyentes. Para corroborar las hipótesis de marcadores se realizó un análisis de Boxplots (figura 84) de la lista de metabolitos de las tablas 34, 35 y 36. Este análisis mostró que sólo hubo dos metabolitos marcadores (4-hydroxi-2-pyrrolidinecarboxylic acid; N-Me y Capsaicina) para la discriminación entre ambas cosechas orgánica y convencional tanto por especies y accesiones. Por otro lado, el análisis Boxplots también corroboró otros metabolitos marcadores específicos para diferenciar cosechas por accesiones. De los resultados obtenidos se observa que no existen grandes diferencias a nivel de metabolitos entre las cosechas orgánica y convencional, solo existen diferencias entre niveles de metabolitos (concentración de metabolitos) y en general ambas cosechas se pueden diferenciar por sus niveles de Capsaicina ya que es el metabolito mayoritario entre los ajíes nativos estudiados (con mayor área relativa), su análisis por LCMS sería mas adecuado porque hay disponibilidad de estándares de Capsaicina y sería un buen marcador para diferenciar la cosecha orgánica de la convencional.

8.3. Metabolómica de las cinco accesiones de ajíes nativos, biomarcadores

Un biomarcador basado en metabolitos es un metabolito que permite delimitar las accesiones o especies. Teóricamente existe una relación estricta entre el genotipo y el fenotipo. Los frutos de ajíes presentan diferentes formas y colores. A esto se le

denomina la plasticidad fenotípica. La expresión del genoma difiere en función de las condiciones ambientales y procesos de adaptación que conducen a diferencias morfológicas. En el genotipo existen los marcadores genéticos neutros que en principio no se ven afectados por las condiciones ambientales (marcadores no involucrados en el proceso de selección). Su expresión en forma de metabolitos deben ser informativos sobre las barreras de las especies y accesiones e independientes de la heterogeneidad y las condiciones ambientales del lugar y distribución de la especie (156). Estos metabolitos biomarcadores podrían provenir de marcadores genéticos neutros.

De acuerdo con los resultados obtenidos, el análisis metabolómico untargeted condujo a la identificación de los biomarcadores que permitieron discriminar las 5 accesiones de ajés nativos. Pese a las diferencias existentes entre las cosechas orgánicas y convencionales el método propuesto en este estudio fue capaz de predecir la misma identidad en los dos tipos de cultivos.

Para hallar los biomarcadores, luego de la adquisición de la data cromatográfica y espectros de masas MS2 por UHPLC-HRMS, el grueso de la información fue procesada previamente con el objetivo de proporcionar datos estructurados en un formato apropiado para el análisis de datos, puesto que los análisis multivariados generalmente se realizan con los resultados de las características metabólicas a una concentración dada (áreas relativas) en lugar de realizarse directamente con los cromatogramas. Una etapa importante de la metabolómica es la identificación de

los metabolitos. En la identificación de los metabolitos, se debe mencionar que existen cuatro niveles de identificación: 1) metabolito identificado, descrito por tener a disposición dos técnicas analíticas ortogonales aplicadas al análisis tanto del metabolito de interés y a un compuesto químico de referencia estándar de presunta equivalencia estructural, con todos los análisis realizados bajo condiciones analíticas idénticas dentro del mismo laboratorio. Ejemplos de datos ortogonales apropiados: masa exacta a través de MS con tiempo de retención; MS de masa exacta y datos de fragmentación o patrón isotópico; Espectros de RMN 2D; espectros de RMN ¹H y/o ¹³C completos, etc. 2) compuesto putativamente anotado, que incluyen espectros de (RMN y/o MS) similares con espectros de librerías públicas o comerciales, 3) clases de compuestos putativamente caracterizados, compuesto desconocido con propiedades espectrales y/o fisicoquímicas consistentes con un particular clase de compuestos orgánico, 4) Un compuesto desconocido, con una señal espectral discernible (RMN, MS u otra) que puede ser detectada, cuantificada y reproducible (21). En este trabajo, la identificación de los metabolitos alcanzó el nivel 2.

El análisis metabolómico fue realizado con la data cromatográfica de las cinco accesiones cultivadas en las tres localidades de Chíncha, Chiclayo y Huaral, ya que un mayor número de muestras puede reducir en cierta medida el efecto de los errores aleatorios y sistemáticos. Además, en el análisis se incluyeron muestras de control de Calidad o QC. El QC es una agrupación de pequeñas alícuotas de cada muestra mezcladas completamente, teóricamente son idénticas, con una

composición metabólica y de matriz de muestra similar a la de las muestras en estudio. El QC proporciona una medida de la repetibilidad y robustez del análisis (157). El diseño del QC permite monitorear el sistema y la estabilidad de la muestra en el transcurso del análisis (158). Los resultados del análisis multivariado por PCA permitieron clasificar las categorías. Una primera categoría fueron las especies. De las cinco accesiones, de acuerdo con la información disponible de la taxonomía entregada por el INIA hay tres especies, *C. baccatum*, *C. chinense* y *C. frutescens*. Los resultados del PCA por especies muestran el clusterizado de las tres categorías, lo que corrobora las observaciones taxonómicas. Además, se observaron las muestras QC en el centro del diagrama indicando que el análisis tuvo éxito.

Los análisis preliminares del modelo OPLS-DA y S-plot de la data cromatográfica general normalizada a escala pareto mostró como biomarcadores putativos de la especie *C. frutescens* a la Capsaicina y Dihidrocapsaicina, de *C. chinense* al S-Malato y de *C. baccatum* al ácido cítrico. Los diagramas boxplots de estos metabolitos corroboran estas observaciones. Sin embargo, la abundancia de la Capsaicina, Dihidrocapsaicina y S-malato expresada en áreas relativas son aproximadamente equivalentes y por lo tanto no se pueden considerar como biomarcadores. De acuerdo a los diagramas boxplots, el ácido cítrico se podría considerar como biomarcador para la especie *C. baccatum* ya que está presente con una alta área relativa, mientras que para las demás especies es baja.

Del PCA también se observó la variación intraclase entre las muestras de las especies debido a la calidad del clusterizado. Los experimentos de metabolómica están inevitablemente sujetos a un componente de variación no deseada, debido a factores como los efectos de batch, largos tiempos de adquisición de data, y variaciones biológicas confusas. Aunque la eliminación de esta variación no deseada es un paso vital en el análisis de los datos de metabolómica, en la actualidad es un “área gris” en la que existe una necesidad reconocida de desarrollar una mejor comprensión de los procedimientos y métodos estadísticos necesarios para lograr resultados biológicos óptimos estadísticamente relevantes (159). Las variaciones se expresan en los niveles de expresión de los metabolitos. Los cinco ajíes nativos tienen como centro de origen la selva central del Perú. Estos fueron llevados a la Costa del Perú para estudios agronómicos de rendimiento, lo que conlleva a los ajíes a pasar por un proceso de adaptación y estrés. Al haber pasado por un proceso de estrés, los ajíes crean variantes genéticas de adaptación. Estas variantes genéticas están directamente relacionadas con el genotipo o metabolitos secundarios. El manejo agrícola de los ajíes fue orgánico y convencional, las características del suelo, el clima, entre otros, son parámetros aleatorios que modifican los niveles del perfil metabólico. Además, de acuerdo a la técnica empleada UHPLC-HRMS el perfil metabólico es una “foto instantánea” de un proceso biológico dinámico que está en continuo desarrollo. Por lo tanto, la metabolómica untargeted de los ajíes nativos es un tema complejo. En este trabajo, la variación intraclase se ha atribuido a los errores aleatorios producidos por el manejo agrícola de las muestras, características del suelo, el clima, etc.

Para eliminar esta variación, se procedió tentativamente con la normalización de la data y corrección de la señal basado en el QC. La normalización de la data se realizó por el método EigenMS. La normalización es un paso importante en el análisis de datos de MS, pero se complica por la alta complejidad de los sesgos. EigenMS ha demostrado que elimina los sesgos de complejidad arbitraria de datos de proteómica y metabolómica (118).

La normalización EigenMS por especies permitió encontrar 93 errores sistemáticos y aleatorios de complejidad desconocida (características o metabolitos) que fueron eliminados de la data original. De igual manera, en la normalización por accesiones se encontraron 132 errores aleatorios y sistemáticos de complejidad desconocida (características o metabolitos) que fueron también eliminados de la data original.

Los análisis por PCA y OPLS-DA de la data cromatográfica normalizada muestran un excelente clusterizado y clasificación categórica de las accesiones y especies. Los resultados del PCA y OPLS-DA muestran valores de calidad de ajuste y de predicción R^2Y y $Q^2Y > 0.9$, que denotan una excelente calidad de los modelos. Todos los modelos fueron validados por CV-ANOVA ($p < 0.05$). En ese sentido, se ha tenido éxito en la eliminación de la variación no deseada.

De los modelos de PCA y OPLS-DA normalizados se extrajo la data normalizada de Chincha para trabajarla separadamente. Un análisis supervisado de OPLS-DA y

análisis S-plot permitió poner en evidencia los biomarcadores putativos por especies y accesiones. De las especies y accesiones se han considerado 8 biomarcadores.

La identificación de biomarcadores por accesiones y especies ha puesto en evidencia las clases químicas involucradas. Del total de metabolitos encontrados en los dos modos de ionización, entre las clases químicas de los biomarcadores se han encontrado nucleósidos, cumarinas, flavonoides, terpenoides y alcaloides. De los nucleósidos y las cumarinas solo existen cuatro metabolitos identificados. De los metabolitos biomarcadores se han encontrado para la especie *C. frutescens* un nucleósido y para *C. baccatum*, una cumarina. De la especie *C. chinense* también se encontró un único lignano, un fuerte candidato de biomarcador para esta especie. De acuerdo con el perfil metabólico, la mayor parte de los metabolitos son minoritarios. Sabiendo de antemano los biomarcadores de las especies y accesiones, se podría pasar de la metabolómica untargeted a metabolómica targeted, tomando en cuenta las clases químicas involucradas en los biomarcadores, de los que se han desarrollado métodos de extracción (160–164). Además, para la confirmación de los metabolitos putativos sería necesaria la utilización de estándares para garantizar la identificación de los metabolitos por los métodos metabolómicos (en el presente trabajo se ha llegado hasta el segundo nivel de identificación), los cuales no siempre están disponibles. Puesto que la concentración relativa de los biomarcadores es baja, se podría elevar la concentración de los metabolitos en el extracto. Para elevar la concentración de los metabolitos haría falta realizar la extracción selectiva de los

metabolitos en base a las clases químicas de los biomarcadores, y posteriormente realizar el análisis metabolómico y los análisis estadísticos por OPLS-DA y S-plot. El papel central de la metabolómica es recopilar datos analíticos y convertirlos en conocimiento biológico. La identificación de los metabolitos de interés (biomarcadores) representa un paso clave en este proceso. Sin embargo, esta identificación también constituye uno de los principales cuellos de botella de todos los estudios metabolómicos porque las bases de datos no siempre están disponibles para un instrumento dado. Como la anotación de un pico cromatográfico es a menudo putativa, la identificación desconocida requiere una determinación completa de la estructura de Novo basada en la interpretación de datos de MS y RMN (165). A pesar de la incertidumbre en la identificación putativa, existe una alta probabilidad de que la mayoría de los compuestos identificados como biomarcadores estén adecuadamente caracterizados debido a sus altos valores de scores de sus patrones de similaridad MS2.

Otra herramienta en la metabolómica es la red molecular de similaridad que utiliza los patrones de fragmentación MS2. La red molecular sirve como método de dereplicación, para anotar metabolitos desconocidos que tengan una alta similaridad a los patrones de fragmentación MS2 de moléculas conocidas, entre ellos los biomarcadores. Desafortunadamente algunos de los metabolitos biomarcadores desconocidos no fueron similares a ningún compuesto identificado. Sin embargo, la red molecular ha permitido poner en evidencia algunos metabolitos

desconocidos pertenecientes a grupos de clases químicas, que tentativamente podrían ser identificados (166).

En contraste con otros estudios metabolómicos untargeted sobre el género *Capsicum*, no hay estudios de autenticidad. Tampoco hay trabajos que reporten los biomarcadores que delimiten las especies o accesiones de ajíes. La mayoría de los trabajos se centran en la implementación de la metabolómica untargeted o perfilado basado en la plataforma analítica de ^1H -RMN de protones. Algunas investigaciones se enfocan en la identificación de marcadores que involucran cambios de los perfiles de metabolitos causados por agentes exógenos. Por ejemplo, en el trabajo de Villa-Ruano *et al.* (167) se trabajó con el perfilado metabólico vía ^1H NMR de protones de los ajíes sanos e infectados (*Capsicum annuum* cv. mirasol) con BMCTV (Beet mild curly top virus) para determinar los marcadores químicos asociados al proceso de infección. Para poner en evidencia los marcadores, se utilizó un modelo OPLS-DA para confrontar la data ^1H -RMN de los ajíes sanos e infectados. Con el modelo OPLS-DA, los valores de importancia VIP fueron puestos en evidencia. Se consideraron valores de $\text{VIP} > 1$ como fuertes marcadores. En el trabajo de Macel *et al.* (168) se implementó la metabolómica untargeted basada en la plataforma analítica HPLC acoplada al espectrómetro de masas híbrido LTQ-Orbitrap FTMS. El objetivo del estudio fue la identificación de fuentes naturales de metabolitos secundarios relacionados con la resistencia a trips (*Frankliniella occidentalis*) en 11 accesiones de *C. annuum* y *C. chinense*. Mediante experimentos de ensayos de discos de hojas, utilizando las hojas de las

plantas en las etapas vegetativas y generativas de crecimiento para alimentar a los trips, se logró clasificar las accesiones en dos categorías, las resistentes y susceptibles. Utilizando estas categorías en el modelo PLS-DA se detectaron los metabolitos dímeros dicapsinósidos como los metabolitos de mayor resistencia a los trips.

También, en el trabajo de Florentino-Ramos *et al.* (169) se estudiaron 11 cultivares mexicanos de *C. annuum*, con el objetivo de establecer su huella digital metabólica mediante espectroscopía ¹H- RMN. Los análisis multivariados se basaron en PCA y LDA (análisis discriminante lineal). Este último fue utilizado para clasificar los cultivares de acuerdo a su perfil de metabolitos. Los resultados mostraron una clara diferenciación entre los 11 cultivares, así como la identificación de tres clases químicas más importantes (carbohidratos, aminoácidos, ácidos orgánicos y vitamina C).

Asimismo, en el trabajo de Cervantes-Hernández *et al.* (170) se describe un estudio metabolómico untargeted utilizando la plataforma analítica UHPLC-QTOF-HRMS, para estimar los perfiles globales de metabolitos del pericarpio, la placenta y las semillas de los frutos del ají Tabasco (*Capsicum frutescens*) en su etapa de maduración. El análisis multivariado se basó en PCA. Los resultados de PCA mostraron una clara diferencia entre los órganos de los frutos por formación de cluster de las muestras debido a sus diferentes perfiles metabólicos. El estudio mostró que el método de metabolómica de LC-MS untargeted es una herramienta

poderosa para la identificación putativa de metabolitos secundarios específicos en *Capsicum*.

De acuerdo a los trabajos mencionados, la metabolómica untargeted puede ser aplicada para resolver problemas específicos. Por lo tanto, para el investigador es de vital importancia el dominio de las técnicas estadísticas multivariadas para desarrollar estrategias específicas que conlleven a resolver los objetivos generales-específicos planteados en una investigación en particular.

8.4. Análisis multivariado de la data SPME-GC-MS, ají fresco y liofilizado

En la naturaleza, las plantas emiten una gran diversidad de compuestos orgánicos volátiles (COVs). Estos actúan en interacciones directas e indirectas para proteger a las plantas de diversas formas de ataques bióticos y facilitan la adaptación en su entorno (171). El género *Capsicum* no es la excepción. La gran diversidad genotípica distribuida en diferentes entornos provee una amplia variabilidad de formas en los frutos, así como una gama de COVs. Algunos estudios realizados en frutos de *Capsicum* así lo corroboran (30, 158–162).

La extracción de metabolitos volátiles fue llevada a cabo por microextracción en fase sólida o SPME. El perfil de metabolitos volátiles depende del tipo de fibra SPME utilizada, ya que poseen diferentes soportes y capacidades de absorción de metabolitos volátiles. Una sola fibra SPME puede absorber un determinado grupo

de compuestos de determinada polaridad y peso molecular. Para cubrir un mayor número de metabolitos de las cinco accesiones de ajíes nativos se realizó la extracción utilizando dos fibras SPME, tanto para los ajíes frescos como para los liofilizados. Los ajíes liofilizados presentan un perfil de volátiles y aroma similar a los ajíes frescos, ya que el efecto de la liofilización no altera significativamente el perfilado en comparación con los metabolitos volátiles de las muestras frescas. Esta afirmación se corrobora en estudios comparativos en frutas tales como el Kiwi, especias como el ajo fresco, en hierbas como la albaca y en *Capsicum* como el pimiento de padrón (*Capsicum annuum*), los cuales mostraron perfiles de metabolitos volátiles de sus productos liofilizados muy similares a los de los productos frescos (177–182). También se ha observado un perfilado metabólico más completo al superponer los dos perfiles cromatográficos de los metabolitos volátiles de los ajíes frescos y liofilizados.

En los estudios de metabolómica por GC-MS, la mayoría de las aplicaciones se realizan mediante ionización electrónica. Este método de ionización siempre se realiza con la misma energía (70 eV) y, por lo tanto, proporciona espectros de MS muy reproducibles. Además, la ionización electrónica MS produce un número importante de fragmentos que tienen patrones característicos para cada analito. Por lo tanto, los metabolitos se pueden identificar de manera eficiente en función de sus patrones de fragmentación después de un procedimiento de deconvolución. La identificación de los metabolitos volátiles fue llevada a cabo utilizando el índice de retención o Kovats calculado a partir de los tiempos de retención de una lista de

alcanos C8-C20 que rindieron identificaciones con altos valores de score o Dot product (183).

De la identificación de los metabolitos se ha puesto en evidencia las clases químicas involucradas, entre las que destacan los ésteres (135 metabolitos) tales como el **Isovalerato de etilo** que expresa aromas característicos a manzana, piña, tutti frutti, fruta, dulce, afrutado, entre otros metabolitos con notas a frutas que contribuyen al aroma de los ajíes. Como segundo grupo se encuentran compuestos alifáticos (43 metabolitos) tales como **(3E,7E)-4,8,12-Trimetiltrideca-1,3,7,11-tetraeno** con nota a frutas. Los sesquiterpenos constituyen el tercer grupo de metabolitos (38 metabolitos) de aromas característicos tales como el **beta-cubebeno**, con notas a afrutado, cítrico, fruta, entre otros, y los monoterpenos tales como el **limoneno** con notas a menta, limón, cítricos, naranja, fresco y dulce.

El estudio metabolómico a partir del perfil de metabolitos volátiles de los frutos de ajíes sirvió para clasificar a las muestras por accesiones y especies. De los análisis multivariados de PCA y OPLS-DA se observó una buena clasificación categórica. Por otro lado, en el género *Capsicum* se ha observado muy pocos estudios metabolómicos enfocados en metabolitos volátiles. Se han reportado estudios que ilustran la potencia de los métodos multivariados para el estudio de esta parte del metaboloma. Por ejemplo, en el trabajo de Jang *et al.* (184), se realizó un análisis metabolómico untargeted del pimiento picante (*Capsicum annuum* "CM334") en el que se estudió su desarrollo por 16, 25, 36, 38, 43 y 48 días después de la anthesis

(DPA) ó floración para analizar sus cambios bioquímicos. Para el análisis de volátiles se utilizó la plataforma GC-TOF-MS. El análisis de componentes principales (PCA) y análisis discriminante de mínimos cuadrados parciales (PLS-DA) fueron las herramientas multivariadas para comparar las diferencias de metabolitos durante el desarrollo del pimiento picante, lo que permitió observar 3 patrones de distribución distintos en los cambios de los metabolitos, las expresiones génicas y las actividades antioxidantes al inicio (16-25 DPA), en la etapa de interrupción (36-38 DPA) y etapas posteriores (43-48 DPA). Las diferencias significativas en las etapas de desarrollo se determinaron sobre la base del valor VIP ($VIP > 0.7$) y el valor p ($p < 0.05$) de los mínimos cuadrados parciales PLS-DA. Luego, se analizaron las rutas bioquímicas, en el que se pudo revelar interacciones entre metabolitos y genes en la biosíntesis de aminoácidos, capsaicinoides, cadenas de ácidos grasos y flavonoides. Los cambios en la actividad antioxidante también reflejaron las alteraciones en los metabolitos. A diferencia del estudio de Jang, en el presente estudio se considera que las muestras de ajíes nativos estuvieron en la etapa de interrupción. El trabajo de Jang *et al.*, muestra claramente la potencia de los métodos multivariados para estudiar etapas específicas de desarrollo de los ajíes.

En otro estudio sobre pimiento dulce llevado a cabo por Eggink *et al.* (37) se desarrolló un método para describir los atributos químicos sensoriales de 24 líneas e híbridos de pimiento dulce (*C. annuum*). En el presente estudio no se hizo ningún tratamiento de muestra, en cambio en el estudio de Eggink *et al.* se extrajeron los metabolitos volátiles con tratamiento previo de la muestra utilizando los reactivos

de EDTA–NaOH y CaCl₂. Los metabolitos volátiles son líquidos lipofílicos con bajo peso molecular y alta presión de vapor a temperatura ambiente. Las propiedades físicas de estos compuestos les permiten atravesar libremente las membranas celulares y liberarse en el entorno. Por el tipo de fibra utilizada, los metabolitos extraídos de los ajíes liofilizados son metabolitos polares, mientras que los metabolitos del ají fresco son metabolitos semi polares o apolares. Es probable que la poca volatilidad de algunos metabolitos sea debido a su afinidad con el agua (la humedad de los ajíes es cercana al 80%). A diferencia del estudio realizado por Eggink *et al.*, en el presente estudio se utilizó un baño maría a 60 °C y ultrasonido. En ese sentido es justificable la extracción de metabolitos volátiles de los ajíes liofilizados ya que, al estar exentos de humedad, los metabolitos pueden difundirse más libremente en el espacio de cabeza del sistema.

En el trabajo de Eggink *et al.* los metabolitos fueron extraídos utilizando una fibra SPME. La adquisición de la data, la deconvolución y alineación de los metabolitos fue llevada a cabo utilizando el software MetAlign. Los resultados del trabajo de Eggink *et al.* mostraron la identificación de 224 metabolitos volátiles. En el presente estudio se han identificado 386 metabolitos (ambas fibras SPME, ajíes frescos y liofilizados). Los resultados del análisis jerárquico de cluster (HCA) y PCA utilizando patrones de intensidad de los 224 volátiles mostraron las principales diferencias entre todos los genotipos. Esta matriz de datos fue correlacionada con la información de metabolitos no volátiles y de análisis sensorial para poner en evidencia los descriptores químicos del sabor y aroma utilizando un enfoque de

predicción de Random Forest. También, Eggink *et al.* pudieron contrastar los metabolitos entre los genotipos causados por grupos de compuestos volátiles y no volátiles, que podrían estar relacionados con vías metabólicas y precursores bioquímicos comunes. En comparación al presente estudio metabolómico de ajíes nativos, Eggink *et al.* estudiaron una especie de ají, y el tipo de correlación que utilizaron fue una regresión lineal simple. Para el presente estudio se pone en evidencia la dificultad del análisis ya que se está trabajando con tres especies diferentes, lo que hace necesario la utilización de un modelo de regresión multivariado.

En otro estudio más reciente, realizado por Debnath *et al.* (185), se utilizó un enfoque de metabolómica basada en GC-MS en el que se trabajó con 6 cultivares de ajíes de la especie *C. annuum*, 2 ajíes de la especie *C. frutescens* y 1 cultivar de la especie *C. chinense*. En este estudio se extrajeron los metabolitos volátiles del pericarpio y las semillas de ajíes de una fracción apolar utilizando como solvente hexano, extraídos en diferentes etapas de maduración. Los análisis multivariados utilizados fueron PCA, PLS-DA y s-PLS-DA desarrollados en la plataforma Online MetaboAnalyst. Los metabolitos no polares de las especies de *Capsicum* y sus variedades fueron correlacionados con sus etapas de maduración progresiva. Los resultados mostraron distintos cambios durante la maduración en función de los perfiles de metabolitos. En este trabajo, nuevamente, los métodos multivariados muestran su desempeño para delimitar las especies en función del perfil de metabolitos volátiles.

De acuerdo con estas investigaciones se observa que los métodos multivariados pueden delimitar las especies. El trabajo de Debnath *et al.* es uno de los primeros trabajos que muestra la performance de los métodos multivariados para delimitar las especies de *C. annuum*, *C. frutescens* y *C. chinense*. Los análisis de S-PLSDA para la discriminación de estas especies mostró que las especies *C. frutescens* y *C. annuum* presentan similares patrones de variables correlacionadas, lo que podría interpretarse como la misma especie. Esta discrepancia podría deberse a que solo se han considerado metabolitos apolares, lo que sería una deficiencia del método. En el presente trabajo de ajíes nativos, la utilización de las dos fibras SPME ha cubierto tantos metabolitos polares, semipolares y apolares; obteniéndose de esta manera un perfilado más completo para la delimitación de las especies por metabolómica untargeted basado en metabolitos volátiles. También se utilizó el método supervisado OPLS-DA para evaluar la capacidad predictiva de los modelos que arrojaron un valor de ($Q^2Y > 0.9$), lo que permite decir que los modelos OPLS-DA puede predecir las variables respuesta de las accesiones y especies en base a sus perfiles de metabolitos volátiles.

8.5.Redes Moleculares como herramienta en la identificación de biomarcadores

En las figuras 116 y 117 se observan las dos redes moleculares de ESI (+, -) diseñadas por el software MSDIAL y estructuradas con el software Metgem. Se ha enfatizado el uso de las redes moleculares como herramienta para identificar

biomarcadores. Dentro de cada Nodo (metabolito) se han colocado un gráfico Pie que indica la abundancia del metabolito en las tres especies estudiadas. Un nodo con un solo color indicaría una abundancia del 100% en esa especie y por lo tanto sería un buen candidato de biomarcador. De esta manera se han descubierto biomarcadores analizando las redes moleculares. Por otro lado, las redes pueden servir como herramienta de dereplicación para elucidar potenciales metabolitos biomarcadores via similaridad MS2. Esta herramienta podría permitir elucidar metabolitos no listados en las bases de datos. Sin embargo, este análisis es de ardua labor y se requiere de bastante experiencia para resolver los espectro MS2 y llegar a la estructura de la molécula desconocida, lo cual no es el objetivo del presente estudio.

8.6. Validación del modelo de huella metabólica, biomarcadores

La validación es un paso obligatorio cuando se realiza una investigación exploratoria, y sirve para garantizar que las hipótesis generadas sean correctas y para evitar conclusiones erróneas. El proceso de validación incluye la comprobación del ajuste, así como la verificación de la validez predictiva del modelo aplicándolo a nuevas muestras, y la evaluación de la validez de la interpretación del modelo en términos de significado biológico o químico. La capacidad predictiva de un modelo se estima mediante el estadístico Q^2 que mide qué tan bien se pueden predecir los datos para muestras no utilizadas en la determinación del modelo. Por otro lado, el conjunto de calibración (muestras

utilizadas para construir el modelo) debe cubrir todas las variaciones posibles (máxima varianza) en los datos X para que el modelo sea válido en el rango de las condiciones de los datos. La validación externa es la mejor manera de validar un modelo y consiste en calcular predicciones para nuevas muestras (El conjunto de predicciones no se utiliza para construir el modelo), pero también es el método más exigente debido al tiempo y costo adicional necesario para el análisis de las muestras. Una opción a la validación externa es la validación interna, también conocida como validación cruzada (CV), que se utiliza cuando no se dispone de un conjunto de predicción (191).

El modelo de OPLS-DA generado de la data cromatográfica de los ajíes de la localidad de Chincha permitió la identificación de los biomarcadores cuya finalidad fue la delimitación de las especies y accesiones. A este modelo OPLS-DA se le consideró como conjunto de calibración o de entrenamiento. Es decir, la data cromatográfica de los ajíes de Chincha sirvió para desarrollar el modelo OPLS-DA. Por otro lado, se consideraron dos nuevos conjuntos de trabajo de data cromatográfica, de los ajíes de la localidad de Chiclayo y Huaral con los cuales se construyeron dos modelos OPLS-DA. Estos modelos fueron denominados conjuntos de predicciones. Todos los modelos fueron validados internamente por validación cruzada y CV-ANOVA ($p < 0.05$). Los valores de $R^2Y > 0.9$ y $Q^2Y > 0.9$ de los modelos OPLS-DA generados revelaron excelentes capacidades de ajuste y predicción. Los resultados de probabilidades del análisis por predicción de las muestras de Chincha a partir de las muestras de Chiclayo y Huaral revelaron una excelente capacidad de predicción ($YPredPS > 0.5$). Con la clasificación OPLS-DA

se tuvieron muestras que pertenecen a diferentes clases, ya sean las especies (3 especies) o accesiones (5 accesiones) y para cada clase fue calculado un modelo OPLS-DA local con un conjunto completo de parámetros del modelo. Cada clase tuvo la capacidad de clasificar las observaciones o muestras en el conjunto de predicción por proximidad en uno de los modelos de clase (especies ó accesiones). Conceptualmente, en los modelos de clase para los diferentes tipos de muestras se genera una envolvente alrededor de los puntos de datos, es decir, un tamaño del modelo o borde del modelo (generalmente es definido por la distancia crítica del parámetro D_{ModX}). Los resultados de la evaluación de las distancias de Y predichos por especies y accesiones de las localidades de Chincha, Chiclayo y Huaral, respectivamente mostraron unas distancias que superan levemente el valor crítico 0.05. Las observaciones con un D_{ModX} más grande que D_{Crit} son valores atípicos. Cuando D_{ModX} es dos veces D_{Crit} , son valores atípicos fuertes. Esto indica que estas observaciones son diferentes de las observaciones normales con respecto a la estructura de correlación de las variables. Se observó un solo valor atípico fuerte. En general los valores atípicos son leves y las distancias son muy próximas a los modelos de las clases especies o accesiones (los residuales con respecto a la data de predicción son leves y en promedio se ajustan bien).

Las probabilidades infieren que las clases predecidas (clase accesión y clase especie) evaluadas a partir de los conjuntos de prueba (data de Chiclayo y Huaral) en el conjunto de entrenamiento y da valores de respuesta de scores predecidos caen dentro de la clase de pertenencia de clases en el conjunto de predicción (data de

Chincha) y por lo tanto están correctamente clasificadas por especies o accesiones según sea la evaluación. Por consiguiente, esto nos indicaría que es muy probable que los metabolitos identificados como biomarcadores permanezcan constantes e invariables y mantengan una independencia de las condiciones ambientales, y de los errores aleatorios y sistemáticos.

El método de validación es de gran utilidad para propósitos de autenticidad en alimentos (191,192). El tema fundamental es la disposición de una data de entrenamiento y de predicción. Esta última es indispensable para la validación externa del modelo OPLS-DA.

Entre los estudios que utilizan la metodología de validación se encuentra el trabajo de Novotná *et al.* (193). Dicho trabajo consistió en la clasificación de cultivos orgánicos versus cultivos convencionales de 40 accesiones de tomates (5 variedades) y 24 accesiones de pimiento dulce (3 variedades) (*C. annuum*) a través de un enfoque metabolómico de huella digital untargeted en el que se estableció un modelo PCA+LDA para la clasificación. La plataforma analítica utilizada fue la ionización de análisis directo en tiempo real (DART) acoplado a espectrómetro de masas de tiempo de vuelo (TOFMS), DART-TOFMS. Para el caso de los pimientos se realizó una selección cuidadosa de marcadores potenciales de los espectros DART-MS de los iones con intensidades significativas, 32 marcadores del modo de ionización positivo y 58 marcadores del modo de ionización negativo (cut-off del 2% del ion con la intensidad más alta en el espectro de masas respectivo), con los

cuales se estableció el modelo de PCA + LDA. El modelo LDA, teóricamente se basa en el cálculo de las funciones discriminantes lineales, que maximizan la relación de varianza intergrupo (las clases respectivas) y al mismo tiempo, minimizan la relación de varianza intragrupo de las clases para el reconocimiento de patrones supervisados (194). Para distinguir entre las muestras del sistema de producción orgánico y convencional, los marcadores obtenidos en el modo positivo mostraron mejores resultados. Empleando las técnicas PCA y LDA, la capacidad de reconocimiento lograda fue del 100% y la capacidad de predicción fue del 87.5%. Para lograr estos resultados, se utilizaron 21 componentes principales.

En este estudio, el modelo PCA+LDA permitió clasificar las muestras de los dos sistemas de producción. En LDA se utiliza en un espacio que está definido por un conjunto básico de vectores llamados discriminantes lineales (LD) que son similares a las PC (componentes principales) obtenidas del análisis de PCA. Al igual que las PC, las LD son combinaciones lineales de las variables originales que también son ortogonales entre sí. Sin embargo, se determinan utilizando un criterio bastante diferente, donde la relación entre la variabilidad intraclase y la variabilidad interclase en los datos de calibración se maximizan (194). No obstante, el modelo OPLS-DA empleado en el presente estudio de clasificación de ajíes nativos es un método más potente para la clasificación categórica. El número de componentes principales empleados (dos componentes) para construir el modelo es mucho menor que el modelo PCA+LDA (21 componentes), por lo que el modelo OPLS-DA captura la mayor variabilidad de la data. Además, el modelo OPLS-DA es una

extensión de PLS-DA y este último a su vez se le puede considerar como una versión combinada del método de regresión de mínimos cuadrados parciales (PLS) y LDA (195). Asimismo, cabe señalar que a medida que el tamaño de las muestras aumenta, la tasa de error de clasificación se reduce para los modelos de PLS-DA y OPLS-DA; mientras que con PCA + LDA son inconsistentes con tamaños de muestras grandes. Por lo tanto, PLS-DA y OPLS-DA son más recomendables para tamaños de muestras grandes que la combinación de PCA + LDA en la reducción y clasificación. Se puede considerar que PLS-DA y OPLS-DA son técnicas confiables cuando se trata de grandes conjuntos de muestras para la tarea de clasificación (196). Además, OPLS-DA es una modificación de PLS desarrollada para conjuntos de datos altamente no correlacionados, que elimina la variación de las variables X no correlacionadas con las variables Y. En ese sentido, la interpretación final de los resultados es más fácil debido a la complejidad reducida del modelo final y generalmente mejora la capacidad de predicción del modelo (197). En términos generales el uso de LDA u otros métodos de clasificación requiere que el tamaño de la muestra dentro de las clases sea lo suficientemente grande como para permitir una clasificación adecuada de las muestras (192). En el presente estudio de ajíes nativos peruanos se utilizaron 60 muestras de extractos de ajíes de las tres localidades de Chincha, Chiclayo y Huaral para construir el modelo OPLS-DA. Esto permitió tener un mayor número de muestras para cada clase representativa sea especies ó accesiones en la clasificación categórica. Otra ventaja del presente trabajo en ajíes nativos es la plataforma analítica empleada UHPLC-

HRMS, la cual tiene la ventaja de la separación cromatográfica, lo que mejora la interpretación, contrario a utilizar solo DART-TOFMS.

En otro estudio, realizado por Cubero-Leon *et al.* (198), se aplicó un enfoque de metabolómica de huella digital untargeted comparativa a gran escala para investigar el efecto del sistema de producción agronómica (orgánica y convencional) en la composición de metabolitos de zanahorias y para construir modelos estadísticos con fines de predicción. Se trabajaron un total de 140 zanahorias en cuatro años consecutivos de producción. En la parte analítica, se realizaron diez réplicas biológicas de cada campo que se analizaron por UHPLC-MS en un sistema LC UltiMate 3000 Dionex acoplado a un espectrómetro de masas Bruker micrOTOF II. Se aplicaron con éxito los modelos OPLS-DA para predecir el origen del sistema agrícola de las zanahorias cosechadas en función de las características determinadas. Los conjuntos de datos se exportaron al software estadístico SIMCA-P para el análisis de datos multivariados. Todas las variables (características moleculares) se escalaron a escala Pareto. El análisis de componentes principales (PCA) fue realizado inicialmente para el análisis multivariado sin supervisión. Para el análisis predictivo, se construyeron proyecciones ortogonales a modelos de análisis discriminante de estructuras latentes (OPLS-DA). El rendimiento de todos los modelos fue evaluado por su variación explicada (R^2X para PCA y R^2Y para OPLS-DA) y su capacidad predictiva (Q^2). Se utilizó una validación cruzada interna de siete veces para determinar los componentes significativos de los modelos y, por lo tanto, minimizar el sobreajuste. Los modelos iniciales de OPLS-

DA fueron construidos sobre conjuntos de entrenamiento que fueron dos tercios de las muestras (seleccionadas al azar) y con características moleculares seleccionadas en función de los valores de importancia de las variables en la proyección (VIP)>

1. La validación externa fue llevada a cabo utilizando el tercio restante de las muestras. Los potenciales candidatos para biomarcadores discriminantes fueron seleccionados utilizando el gráfico S-plot generados por los modelos OPLS-DA.

El trabajo de Cubero Leon tiene muchas similitudes, desde el punto de vista de las técnicas estadísticas y analíticas empleadas, con el presente trabajo de ajíes nativos. Las diferencias radican en los años de producción de las cosechas. En el trabajo de Cubero Leon se trabajó con cosechas de 4 años de producción, mientras que en el proyecto de ajíes nativos se ha trabajado con cosechas de solo un año de producción, que en realidad fueron tres cosechas si se considera que los ajíes fueron cultivados en tres localidades. Este sería quizás un punto débil del proyecto de ajíes nativos. En este proyecto, no se han trabajado con ajíes de otras temporadas debido a cuestiones económicas ya que la producción de ajíes para este estudio proviene de un estudio piloto de accesiones y no variedades de ajíes ya domesticadas. Estas cinco accesiones están todavía bajo estudios agronómicos para su domesticación y masificación. En el proyecto de ajíes se trabajó con una cosecha de producción orgánica y dos cosechas de producción convencional. En ese sentido, no se podría desarrollar un modelo de validación externa para distinguir entre cosechas del tipo orgánica de las convencionales, ya que haría falta tener a disposición de datos de

cosechas de otras temporadas (por lo menos 3 años adicionales) para conseguir un modelo de clasificación sólido y robusto.

Entre otros trabajos que utilizan el mismo enfoque de herramientas estadísticas se puede mencionar el trabajo de Navratilova *et al.* (199) sobre la autenticación del origen geográfico del té verde basada en huellas digitales por UHPLC-HRMS, así como el trabajo de Olga Varrà *et al.* (200) sobre la autenticación de la lubina europea según el método de producción y el origen geográfico mediante una relación de isótopos estables a la luz y análisis de elementos de tierras raras.

Por otro lado, el objetivo del proyecto de ajíes no fue comparar cosechas orgánicas y convencionales sino encontrar los biomarcadores que permitan delimitar las especies y accesiones. Los biomarcadores son metabolitos cuyos niveles se ven afectados por el estrés adaptativo de las plantas. El estrés es originado por las variables aleatorias tales como sistema de producción (agrícola y convencional), lugar de cultivo de los ajíes, tipo de suelo y condiciones climáticas, entre otros. Estas variables fueron superadas con el adecuado tratamiento estadístico que se ha dado a la data de ajíes de esta temporada de producción. Sin embargo, haría falta probar por lo menos con dos temporadas de producción de ajíes adicionales para establecer un modelo de predicción más robusto. Por lo tanto, los resultados obtenidos son preliminares, pero ponen las bases para futuros esfuerzos de completar los estudios de biomarcadores en ajíes nativos.

8.7. Análisis Multibloque, correlaciones

El objetivo de la fusión de datos metabolómicos fue combinar los datos de las dos plataformas metabolómicas con los datos del análisis sensorial medidas en el mismo conjunto de muestras o individuos para encontrar los descriptores químicos del sabor y aroma. Luego de establecer los modelos multivariados OPLS-DA de la data cromatográfica de metabolitos volátiles, no volátiles y data sensorial de los ajíes de Chincha, se procedió a realizar el análisis multibloque con estas tres tablas de datos. El análisis se llevó a cabo utilizando el Package MixOmics del software estadístico de R. Las correlaciones fueron llevadas a cabo utilizando la clasificación categórica por accesiones, ya que los modelos por accesiones presentaron el mejor rendimiento de clasificación. El análisis mostró que cuatro componentes principales capturaban la mayor parte de la información en forma de covarianza según el modelo multivariado de correlaciones por sparse PLS-DA (SPLS-DA). El modelo de correlaciones multibloque fue validado internamente por validación cruzada y criterio de clasificación de distancia de Mahalanobis (p -value = 0.001). Los valores de ajuste de Pearson (R^2) mostraron fuertes correlaciones adyacentes del bloque sensorial frente al bloque de volátiles y no volátiles. Para este análisis se consideró como los valores de respuesta Y a los descriptores del sabor y aroma del bloque sensorial y las variables de medición X a los bloques de metabolitos volátiles y no volátiles. Los resultados mostraron altos valores de correlaciones de Pearson R^2 , lo que indicó que los valores predichos están fuertemente correlacionados con la variable respuesta. La función Circosplot de MixOmics permitió obtener los valores

de las correlaciones. Para discriminar las correlaciones verdaderas de las correlaciones casuales, se utilizó un enfoque Holístico de interpretación. Esta interpretación incluyó los resultados del análisis de AFM, Heatmap y valores VIP del modelo estadístico de análisis sensorial que fueron utilizados como herramientas de soporte para extraer la información de las correlaciones. De esta manera se pusieron en evidencia los descriptores químicos del sabor y aroma. Los análisis CIM y gráficas de individuos fueron gráficos complementarios que reforzaron los resultados del análisis de correlaciones multivariado multibloque. Los gráficos CIM son gráficos no supervisados y permitieron la clasificación de las muestras en función de los tres bloques de datos, mientras que el gráfico de individuos mostró la clasificación categórica por accesiones interbloques del análisis multivariado sparse-PLSDA (S-PLSDA).

Los resultados del análisis multibloque mostraron fuertes valores de correlaciones entre los metabolitos volátiles y no volátiles con las notas del sabor y aroma para cada accesión de ají. Entre los metabolitos volátiles que fueron fuertemente correlacionados se encuentran las clases químicas más importantes del perfil general del metaboloma, tales como los ésteres, alcoholes, monoterpenos y sesquiterpenos.

El ají charapita se caracterizó por sus notas a orégano, cítrico, frutal y maracuyá; estas notas a su vez se caracterizaron por ser ricos en ésteres, alcoholes, cetonas, sesquiterpenos y monoterpenos. Muchos de estos metabolitos mostraron perfiles de

aromas ya conocidos de acuerdo a las bases de datos públicas, tales como notas a menta, limón, cítrico, verde, hierbas, entre otros. Un buen número de metabolitos no tuvieron un perfil de aroma conocido. El 2-methyl-1-penten-3-one y el citronellyl valerate se caracterizaron por sus notas a menta y hierbas. Esto podría justificar la nota a Orégano en el ají charapita. El aroma a maracuyá se puede explicar por el metabolito Isobutirato de bencilo por su nota a tropical. La nota a cítrico puede ser justificada por la presencia de los metabolitos Citral, 2-Undecanone y el 4-Metilpentil 8-metilnonanoato, con aromas a cítrico y naranja, respectivamente. El aroma a frutal se le puede atribuir a los metabolitos Cubeneno, 2-hexen-1-ol y 1-penten-1-ol con sus notas a mango y afrutados, respectivamente. El ají Dulce rojo se caracteriza por sus notas a hierba. Este aroma se puede atribuir a los metabolitos 2-Tridecanone, Butanoic acid, 3-methyl-, 3,7-dimethyl-6-octenyl ester y a gamma-caprolactona por sus notas a herbal y hierbas, respectivamente. El ají miscucho se caracteriza por sus notas a manzana frutal y maracuyá. Estos aromas se pueden atribuir a los metabolitos isovalerato de nonilo y hexil ester del ácido octanoico por sus notas a manzana y afrutado, respectivamente. El ají ayuello se caracteriza por sus notas a hierba y manzana. Estos aromas se pueden atribuir a los metabolitos 3,7-Dimetil-1,3,6-octatrieno, ocimene y al 3-Hexenal por sus notas a hierbas y manzana, respectivamente.

Para el caso del sabor se debe considerar la acción de la saliva humana sobre el sabor y la textura de los alimentos. La saliva puede ejercer diversos efectos sobre el sabor, y como resultado producir cambios en la forma en que se perciben. La

mezcla de la saliva con el pericarpio de los ajíes puede influir en el sabor y causar que las sustancias del gusto y del sabor se diluyan. La α -amilasa, es la enzima presente en la saliva que inicia la digestión de los polisacáridos y desempeña un papel en la descomposición inicial del pericarpio de los ajíes. También, la saliva actúa como un sistema amortiguador, afectando el grado de percepción de la acidez. Además, las proteínas salivales pueden influir en la lubricación y, por lo tanto, posiblemente en la percepción de los atributos como la suavidad y la astringencia (201). También, existe un número de compuestos naturales con capacidad de inhibición de la α -amilasa. Entre ellos se encuentran los flavonoides, como los flavonoles Quercetin (82% de actividad inhibitoria), Quercetrin (45% de actividad inhibitoria), Rutin (40% de actividad inhibitoria) y Kaempferol (40% de actividad inhibitoria entre otros). También los flavonoides del tipo flavonas como Luteolin (100% de actividad inhibitoria) y Luteolin -7-O-glucoside (100% de actividad inhibitoria), entre otros.; las flavononas como la naringenin (26% de actividad inhibitoria); derivados del ácido cinámico como el ácido clorogénico (50% de actividad inhibitoria, 1,4-1,6mM); terpenos como el Lupeol (50% de actividad inhibitoria), entre otros (202).

El ají tomatito rojo se caracteriza por sus notas de sabor a dulce, pimiento y tomate. De acuerdo a las bases de datos públicas, no existen notas que sean atribuibles a los metabolitos correlacionados con estos sabores. Las notas del sabor de tomatito rojo se caracterizan por poseer flavonoides. Los flavonoides son un gran grupo de sustancias naturales con estructuras químicas variables que se biosintetizan en todas

las partes de la planta. Son los pigmentos vegetales más abundantes junto con la clorofila y los carotenoides, que también proporcionan fragancia y sabor a las frutas, flores y semillas, lo que los hace atractivos para insectos, aves y mamíferos, que ayudan a la transmisión de polen o semillas (203). Un tipo de flavonoide son las flavanonas, encontradas exclusivamente en los cítricos en sus formas glucosídicas. Por ejemplo, el jugo de toronja (*Citrus paradisi*) contiene naringina (naringenin-7-Oneohesperidoside) y el jugo de naranja naringina (naringenin-7-O-rutinoside). Las flavanonas también se encuentran en los tomates (una especie de las solanaceas) y ciertas plantas aromáticas como la menta, pero solo en los cítricos están presentes en altas concentraciones (204). En el perfil del sabor de tomatillo rojo se han encontrado nueve flavonoides, entre los que se encuentra Vicenin-2. El Vicenin-2 se ha encontrado en frutas y especias tales como naranja dulce, limones y albahaca dulce. Otro flavonoide es el Edgeworthin; 6-Me éter, 7-Ac que es común en Ginseng. Al resto de flavonoides no se les asocia a un alimento en particular. Este conjunto de metabolitos bien podría representar a las dos notas de sabor, tomate y pimienta. Estos también no tendrían acción inhibitoria sobre la α -amilasa, produciendo la hidrólisis espontánea de los O-glucosidos que producirían el sabor a dulce.

El segundo grupo de metabolitos del sabor de tomatillo rojo son los alcaloides. Los alcaloides se asocian con el sabor amargo. Por ejemplo, en plantas que contienen alcaloides como la cinchona, la belladona o el aconitum el sabor amargo se ha asociado con el sabor de sus extractos, y posteriormente con los compuestos puros

después del aislamiento. Esta asociación es tan fuerte que el sabor amargo de los alimentos todavía se atribuye a veces a la presencia de alcaloides, incluso cuando no se ha demostrado mediante el aislamiento de ningún principio activo y la evaluación sensorial de compuestos individuales. Hay cientos de compuestos amargos de muchas otras clases químicas, tales como los terpenos, aminoácidos y flavonoides. En los alimentos, muchos principios amargos tienen una estructura no alcaloide; por ejemplo, el flavonoide naringina, que le da el sabor amargo a la toronja, la humulona, que se encuentra en la cerveza o la lactucina, un sesquiterpeno lactona encontrado en la lechuga, son solo algunos ejemplos. Esta diversidad en la estructura química de los compuestos amargos tiene un paralelo con la gran variabilidad del amargor (205). Entre los alcaloides que contribuyeron al sabor de tomatito rojo se encontraron seis alcaloides. Los alcaloides que son más representativos son el 3-Carboxy-1-methylpyridinium betaine, 4-Hydroxybenzoylcholine y Capsaicina; Análogo de nonanoilo, comunmente encontrados en hierbas, condimentos y especias. Estos metabolitos podrían ser metabolitos que resaltan los sabores a tomate y pimiento o enmascarar los sabores como es el caso de la Capsaicina; análogo de nonanoilo que le confiere a los ajíes la sensación a picante.

El tercer grupo de metabolitos encontrados en el sabor del ají tomatito rojo son los carbohidratos. Los carbohidratos naturales incluyen azúcares naturales, alcoholes de azúcar, glucósidos terpenoides, así como algunos aminoácidos y polifenoles. Los compuestos de sabor dulce cubren un gran diversidad química y estructural. Entre

algunos ejemplos representativos se encuentran: la glicirricina, esteviósido, Taxifolin 3-O-acetate, monatina y brazzeína. Una estructura muy notable es la brazzeína que es una proteína dulce que se aisló del fruto de la planta trepadora de África occidental Oubli (*Pentadiplandra brazzeana* Baillon). En términos generales, la brazzeína es 500 veces más dulce que la sacarosa en comparación con la solución de sacarosa al 10% (206). En el sabor de tomatito rojo se han encontrado cinco derivados de carbohidratos, que podrían explicar el sabor dulce. La enzima α -amilasa podría ser responsable de la descomposición de los carbohidratos en azúcares simples. El resto de metabolitos con clases químicas de terpenoide, esteroides y aromáticos podrían contribuir al sabor dulce debido a que sus agliconas se encuentran unidas a moléculas de azúcar tal como la glucopiranososa que podrían ser hidrolizadas por la α -amilasa y producir el sabor dulce. También, los aminoácidos Pentosa + prolina y Tirosina; O- β -D-glucopiranosido podrían ser hidrolizados y liberar los azúcares pentosa y glucosa para contribuir con este atributo.

De la misma manera se puede examinar el sabor del ají dulce rojo. Este se caracteriza principalmente por sus notas de dulce y pimienta. El número de clases químicas para estos dos sabores son: cinco flavonoides, cinco terpenoides, dos aminoácidos, cinco compuestos aromáticos, nueve alcaloides y cinco esteroides. Los flavonoides, están presentes en gran parte de las frutas frescas y en particular podrían contribuir con la nota de pimienta en el ají dulce rojo. Los alcaloides de sabor “amargo”, también podrían contribuir para realzar ó enmascarar las notas a

pimiento o dulce. Al parecer, los flavonoides y alcaloides no ejercen acción inhibitoria sobre la enzima α -amilasa ya que la nota del sabor dulce está presente. El resto de metabolitos podrían contribuir al sabor dulce ya que sus agliconas se encuentran unidas a O-glucósidos y por consiguiente ser hidrolizadas hasta sus azúcares más simples y producir el sabor dulce.

Para el ají Ayucello, su nota de sabor más importante es la acidez. Las clases químicas con sabor ácido son: ocho esteroides, un alcaloide, tres compuestos aromáticos, tres flavonoides y seis terpenoides. Para los flavonoides, por ejemplo, la acidez es mayor para fenoles sustituidos que sin sustituir. Asimismo, la acidez de los flavonoides disminuye con la formación de enlaces de hidrógeno intramoleculares (207). En ese sentido, el metabolito 3,4',5,7,8-Pentahidroxi-6-metilflavona; 3-Me éter, 4',8-di-O- β -D-glucopiranosido y los otros dos flavonoides encontrados en el sabor ácido del ají ayucello, podría explicar las notas de acidez. También, de los ocho esteroides hay algunos que contienen en su estructura hasta cuatro grupos hidroxilo como el esteroide 14,17,20,27-Tetrahydroxy-1-oxowitha-3,5,24-trienolide; (14 α ,17 β ,20S,22R)-forma, que podría atribuirse por analogía a los flavonoides el sabor ácido al ají ayucello. El compuesto aromático 3,4-Dihydroxybenzoic acid; 3-Me ether, 4-O-[3,4-dihydroxy-E-cinnamoyl-(\rightarrow 6)- β -D-glucopyranoside], un ácido aromático, también podría explicar la acidez del ají ayucello. Los aminoácidos también podrían contribuir a la acidez. El aminoácido Glutamilleucina, muy común en especias como el ajo, podría explicar la nota de acidez. El resto de metabolitos, terpenoides, alcaloides y esteroides podrían inhibir

la acción de la enzima α -amilasa, lo que explicaría las bajas notas de percepción a dulce en el ají ayucillo.

Los resultados muestran que el análisis de datos y la técnica de ensamblaje de tablas de datos de diversas fuentes ómicas para el estudio de propiedades sobre un mismo conjunto de muestras ofrecen una mejor comprensión del sabor y aroma de los ajíes. Este sería el primer estudio de análisis multibloque empleando las tres tablas de datos involucradas para elucidar los descriptores químicos del sabor y aroma. Dado, que es la primera vez que se reportan estos resultados, no existen resultados de comparación con estos datos.

Los trabajos referidos a la unión de tablas de datos de diversas fuentes ómicas son muy escasos, debido a que esta metodología es novedosa. Como consecuencia de ello, hasta la fecha solo existen ejemplos muy concretos de aplicación que demuestran la potencia del método estadístico.

Se puede citar el trabajo de Zengin *et al.* (208) que trabajó con *Ocimum americanum* L. o Albahaca canosa (Lamiaceae), un condimento alimentario común y también usado en la medicina tradicional en el tratamiento de varias enfermedades humanas. El objetivo del trabajo de Zengin fue delinear los perfiles biológicos y fitoquímicos de las hojas y flores preparadas por diferentes solventes extractivos (acetato de etilo, metanol y agua). Los extractos de agua de flores y hojas mostraron un fuerte potencial reductor y de eliminación de radicales. Tanto los extractos de MeOH

como de acetato de etilo de las hojas y las flores fueron capaces de inhibir la acetilcolinesterasa, la butirilcolinesterasa y la tirosinasa. Solo el extracto de flor de acetato de etilo mostró efectos antioxidantes.

Para el análisis estadístico utilizaron el paquete MixOmics R y el software Xlstat 2017. Se presentaron resultados del análisis supervisado por S-PLSDA en gráficos de individuos de las flores y las hojas en los tres solventes utilizados, gráficos de Mapa de Imagen Agrupada o CIM, gráficos de correlaciones interbloques de las tres tablas de datos: compuestos bioactivos, habilidades antioxidantes y actividades enzimáticas inhibitorias, que mostraron fuertes correlaciones interbloques con coeficientes de ajuste ($R^2 \geq 0.9$). También evaluaron la relación entre los compuestos bioactivos y las actividades biológicas estudiadas resultantes del coeficiente de correlación ($r = 0.70$) mediante el gráfico Circosplot. Finalmente, los datos se analizaron utilizando ANOVA de dos vías para investigar las diferencias entre las muestras, con $p < 0.05$ que se consideró estadísticamente significativo y análisis Post-Hoc múltiple de Tukey.

En otro ejemplo de aplicación, se encuentra el trabajo de Yin *et al.* (209) donde se estudió la expresión radiológica integrativa para predecir subtipos moleculares de carcinoma primario de células renales de células claras. En este trabajo se aplicó la misma estrategia de análisis multibloque.

8.8.Perspectivas Metabolómicas con enfoque hacia metabolitos desconocidos

Se ha observado que el 49% de los metabolitos son desconocidos y por lo tanto existe una gran base de datos de metabolitos aún por elucidar. El descubrimiento de los metabolitos dependerá del target del problema de investigación. En ese sentido se podría proponer un estudio de actividad anticáncer, actividad antioxidante ó actividad antibacteriana (anti *Helicobacter pylori*) para el extracto crudo de *Capsicum*. Algunos antecedentes de estudios sobre extractos de ajíes se refleja en el estudio de Yumiko Hirokawa *et al* con extractos de ajíes de *C. chinense* en donde se observan resultados con actividad anticáncer de mama y próstata (186). También se reporta la actividad antioxidante de los frutos del género *Capsicum* (187). Estudios etnobotánicos muestran el uso del fruto de *Capsicum chinense* como remedio estomacal, antídoto para el dolor abdominal y para contrarrestar la intoxicación por bebidas alucinógenas (188). Puesto que los ajíes nativos son de origen silvestre, estos podrían contener metabolitos interesantes con las propiedades medicinales mencionadas, ya que los ajíes nativos al haber pasado por un proceso de adaptación diferente a sus contrapartes domesticados posiblemente hallan desarrollado genes de resistencia a plagas y parásitos entre otros y por consiguiente metabolitos secundarios especializados. Por tal motivo sería interesante realizar un screening para revisar sus potenciales actividades biológicas. En función de los resultados se podría realizar un aislamiento bioguiado mediante cromatografía de media presión o MPLC y con la ayuda de metabolómica untargeted por UHPLC-HRMS, análisis multivariado y redes moleculares analizar

cada fracción obtenida y realizar las anotaciones de los compuestos activos. Con las actividades obtenidas se podría realizar un modelo de correlación OPLS de perfiles de los extractos vs actividad biológica para encontrar los metabolitos activos dentro de cada extracto. Finalmente, mediante análisis del extracto más activo y cromatografía UHPLC-RMN se podría elucidar el compuesto activo (189). Ejemplos de aplicaciones del enfoque metabolómico para elucidar compuestos actividades específicos de otras plantas se encuentra el trabajo de Pedro Vasquez *et al* en donde se identificaron potenciales metabolitos con actividad antiprotozoaria de las plantas del género *Piper* por correlación OPLS y utilizando herramientas de redes moleculares (190). De igual manera se ha aplicado la misma estrategia para la dereplicación de metabolitos con actividad antioxidante mediante análisis UHPLC-HRMS y redes moleculares para identificar metabolitos con propiedades con actividad redox por evaluación de la capacidad de las moléculas para reducir el DPPH (2,2-diphenyl-1-picrylhydrazyl) y los radicales superóxido ($O_2\cdot^-$) utilizando espectroscopia de resonancia de espín electrónico y UV-Vis (ESR), respectivamente (166).

9. CONCLUSIONES

- El perfilado de metabolitos volátiles por SPME-GCMS de los frutos frescos y liofilizados de las cinco accesiones de ajíes nativos de la localidad de Chíncha puso en evidencia las principales clases químicas de los ajíes de las cuales destacan los ésteres, monoterpenos y sesquiterpenos.
- El perfilado de metabolitos no volátiles por UHPLC-HRMS Orbitrap de los frutos enteros liofilizados de las cinco accesiones de ajíes nativos de la localidad de Chíncha mostró la presencia de las principales clases químicas de los ajíes de las cuales destacan los terpenoides, flavonoides y alcaloides.
- Mediante el modelo estadístico de huella metabólica generado a partir de la data cromatográfica de metabolitos no volátiles de los ajíes nativos de la localidad de Chíncha se han determinado los principales biomarcadores que permiten la discriminación de las especies y accesiones; estos biomarcadores incluyen compuestos tipo flavonoides, cumarinas, lignanos y terpenoides.
- La validación externa del modelo estadístico de huella metabólica de los frutos de ajíes procedentes de la localidad de Chíncha confrontada con los modelos estadísticos de los frutos procedentes de las localidades de Chiclayo y Huaral permitió asegurar la correcta clasificación de los ajíes por especies y accesiones y la continuidad de los metabolitos biomarcadores en cualquier espacio de cultivo.
- Mediante el enfoque de análisis supervisado S-PLSDA de correlación multibloque se pudo predecir los metabolitos volátiles y no volátiles vinculados a los atributos sensoriales de las cinco accesiones de ajíes de Chíncha. Los

metabolitos del aroma identificados pertenecen a las clases químicas de ésteres, monoterpenos y sesquiterpenos; los cuales presentaron notas a manzana, frutas, tropical, hierbas y menta. Por otro lado, los metabolitos del sabor a tomate, pimiento y dulce fueron vinculados a las clases químicas de flavonoides, terpenoides, alcaloides, aromáticos y esteroides.

10.RECOMENDACIONES

- Se recomienda utilizar un mayor número de panelistas para el análisis sensorial de los ajíes y un mayor número de réplicas (4 réplicas por ají). Esto permitirá tener a disposición un tamaño muestral que sea más robusto, logrando disminuir la influencia de valores atípicos de scores causados por percepciones subjetivas.
- Se recomienda utilizar otro sistema de solvente para la extracción de los metabolitos no volátiles en el análisis por UHPLC-HRMS, como por ejemplo acetato de etilo.
- Se recomienda realizar el análisis de metabolitos volátiles con un equipo GC-HRMS, para mejorar la precisión de las masas y la identificación de los metabolitos volátiles.
- Se recomienda analizar los ajíes frescos con la fibra SPME DVB/CAR/PDMS (fibra ploma) para el análisis de metabolitos volátiles polares.
- Se recomienda realizar un diseño experimental para la extracción de metabolitos volátiles por SPME-GCMS utilizando software de diseño experimental como Design Expert.
- Se recomienda repetir los experimentos con otras cosechas de ajíes de otras temporadas, por lo menos dos años consecutivos.

12. BIBLIOGRAFÍA

1. Ortiz R, Delgado de la Flor F, Alvarado G, Crossa J. Classifying vegetable genetic resources—A case study with domesticated *Capsicum* spp. *Sci Hortic*. 13 sept 2010;126(2):186-91.
2. Attokaran M. Natural food flavors and colorants. 2nd edition. Hoboken: Wiley-Blackwell; 2017. (Institute of food technologists' series).
3. Wildman REC, éditeur. Handbook of nutraceuticals and functional foods. 2nd ed. Boca Raton: CRC/Taylor&Francis; 2007. 541 p.
4. Edwards SE, Rocha I da C, Williamson EM, Heinrich M. Phytopharmacy: An Evidence-Based Guide to Herbal Medical Products [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2015 [cité 1 mars 2019]. Disponible sur: <http://doi.wiley.com/10.1002/9781118543436>
5. Amit KD. *Capsicum*: The genus *Capsicum* [Internet]. CRC Press; 2004 [cité 28 févr 2019]. Disponible sur: <https://www.taylorfrancis.com/books/9780203381151>
6. Walsh BM, Hoot SB. Phylogenetic Relationships of *Capsicum* (Solanaceae) Using DNA Sequences from Two Noncoding Regions: The Chloroplast *atpB-rbcL* Spacer Region and Nuclear waxy Introns. *Int J Plant Sci*. 1 nov 2001;162(6):1409-18.
7. Hernández-Verdugo S, Dávila P, Oyama K. Síntesis del conocimiento taxonómico, origen y domesticación del género *Capsicum*. 2017. 65 p.
8. Eshbaugh WH. The genus *Capsicum* (Solanaceae) in Africa. *Bothalia* Vol 14 No 34 1983 [Internet]. 1983; Disponible sur: <https://abcjournal.org/index.php/abc/article/view/1252>
9. Bosland PW, Votava EJ. Peppers: vegetable and spice capsicums. 2nd ed. Cambridge, MA: CABI; 2012. (Crop production science in horticulture series).

10. Pickersgill B. Genetic resources and breeding of *Capsicum* spp. *Euphytica*. 1 juill 1997;96(1):129-33.
11. Alexander DC, Levett PN, Turenne CY. Chapter 21 - Molecular Taxonomy. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J, éditeurs. *Molecular Medical Microbiology (Second Edition)* [Internet]. Boston: Academic Press; 2015. p. 369-79. Disponible sur: <http://www.sciencedirect.com/science/article/pii/B9780123971692000214>
12. Cicconardi F, Fanciulli PP, Emerson BC. Collembola, the biological species concept and the underestimation of global species richness. *Mol Ecol*. nov 2013;22(21):5382-96.
13. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, et al. Cryptic species as a window on diversity and conservation. *Trends Ecol Evol*. 1 mars 2007;22(3):148-55.
14. Porco D, Potapov M, Bedos A, Busmachiu G, Weiner WM, Hamra-Kroua S, et al. Cryptic Diversity in the Ubiquist Species *Parisotoma notabilis* (Collembola, Isotomidae): A Long-Used Chimeric Species? Steinke D, éditeur. *PLoS ONE*. 26 sept 2012;7(9):e46056.
15. KARP A, BUIATTI M, SEBERG O. Molecular Techniques in the Assessment of Botanical Diversity. *Ann Bot*. 1 août 1996;78(2):143-9.
16. Henry RJ. *Practical applications of plant molecular biology*. London: Chapman & Hall; 1997. 258 p.
17. De Masi L, Castaldo D, Galano G, Minasi P, Laratta B. Genotyping of fig (*Ficus carica* L) via RAPD markers. Vol. 85. 2005. 2235 p.
18. Lengua Cabrera RG. *Caracterización Molecular de las colecciones Nacionales de Ajíes Capsicum spp.* Del Instituto Nacional de Innovación agraria (INIA) y la Universidad Nacional agraria La Molina (UNALM) Mediante Marcadores Moleculares [Tesis para optar el título de Biólogo]. Universidad Nacional Agraria La Molina; 2018.

19. Moscone EA, Scaldaferrero M, Grabielle M, Cecchini N, Sánchez García Y, Jarret R, et al. The evolution of Chili Peppers (*Capsicum-Solanaceae*): a cytogenetic perspective. 2014.
20. Toquica SP, Rodríguez F, Martínez E, Duque MC, Tohme J. Molecular Characterization by AFLPs of *Capsicum* Germplasm from the Amazon Department in Colombia. *Kluwer Academic Publishers*. 2003;50(6):639-47.
21. Viant MR, Kurland IJ, Jones MR, Dunn WB. How close are we to complete annotation of metabolomes? *Curr Opin Chem Biol*. févr 2017;36:64-9.
22. Jacob M, Malkawi A, Albast N, Al Bougha S, Lopata A, Dasouki M, et al. A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. *Anal Chim Acta*. 26 sept 2018;1025:141-53.
23. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*. 22 mars 2012;13:263.
24. Dudley E, Yousef M, Wang Y, Griffiths WJ. Chapter 2 - Targeted metabolomics and mass spectrometry. In: Donev R, éditeur. *Advances in Protein Chemistry and Structural Biology* [Internet]. Academic Press; 2010. p. 45-83. Disponible sur: <http://www.sciencedirect.com/science/article/pii/B9780123812643000023>
25. Zhou J, Yin Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst*. 2016;141(23):6362-73.
26. Wolfender J-L, Glauser G, Boccard J, Rudaz S. MS-based plant metabolomic approaches for biomarker discovery. Vol. 4. 2009. 1417 p.
27. Worley B, Powers R. Multivariate Analysis in Metabolomics. *Curr Metabolomics*. 1 nov 2012;1(1):92-107.

28. Jing L, Lei Z, Zhang G, Pilon AC, Huhman DV, Xie R, et al. Metabolite profiles of essential oils in citrus peels and their taxonomic implications. *Metabolomics*. 1 août 2015;11(4):952-63.
29. Wahyuni Y, Ballester A-R, Sudarmonowati E, Bino RJ, Bovy AG. Metabolite biodiversity in pepper (*Capsicum*) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. *Phytochemistry*. août 2011;72(11-12):1358-70.
30. Wahyuni Y, Ballester A-R, Tikunov Y, de Vos RCH, Pelgrom KTB, Maharijaya A, et al. Metabolomics and molecular marker analysis to explore pepper (*Capsicum* sp.) biodiversity. *Metabolomics*. févr 2013;9(1):130-44.
31. Ministerio del Ambiente P. CIISB | Centro de Intercambio de Información Sobre Seguridad de la Biotecnología del Perú | Catálogo de pasaporte del germoplasma de *Capsicum* [Internet]. 2015 [cité 22 mars 2019]. Disponible sur: http://bioseguridad.minam.gob.pe/publicaciones_notas/catalogo-pasaporte-coleccion-germoplasma-capsicum/
32. INIA. INIA y Universidad Agraria consolidarán Colección Nacional de Germoplasma de ajíes peruanos – Instituto Nacional de Innovación Agraria [Internet]. 2018 [cité 22 mars 2019]. Disponible sur: <http://www.inia.gob.pe/2018-nota-128/>
33. Hootman RC, éditeur. *Manual on descriptive analysis testing for sensory evaluation*. Philadelphia, PA: ASTM; 1992. 52 p. (ASTM manual series).
34. McCLOSKEY LP, Sylvan M, Arrhenius SP. Descriptive analysis for wine quality experts determining appellations by chardonnay wine aroma. *J Sens Stud*. avr 1996;11(1):49-67.
35. Menezes AGT, Batista NN, Ramos CL, Silva AR de A e, Efraim P, Pinheiro ACM, et al. Investigation of chocolate produced from four different Brazilian varieties of cocoa (

Theobroma cacao L.) inoculated with *Saccharomyces cerevisiae*. *Food Res Int.* mars 2016;81:83-90.

36. Juárez-Barrientos JM, de Montserrat Tejada-Paz M, de Jesús Ramírez-Rivera E, Aguirre-Cruz A, Rodríguez-Miranda J, Martínez-Sánchez CE, et al. Use of quantitative descriptive analysis (QDA) coupled with multivariate statistical methods to detection and discrimination of adulterated fresh cheeses. *J Sens Stud.* févr 2019;34(1):e12479.

37. Eggink PM, Maliepaard C, Tikunov Y, Haanstra JPW, Bovy AG, Visser RGF. A taste of sweet pepper: Volatile and non-volatile chemical composition of fresh sweet pepper (*Capsicum annuum*) in relation to sensory evaluation of taste. *Food Chem.* mai 2012;132(1):301-10.

38. De Masi L, Siviero P, Castaldo D, Cautela D, Esposito C, Laratta B. Agronomic, chemical and genetic profiles of hot peppers (*Capsicum annuum* ssp.). *Mol Nutr Food Res.* 1 août 2007;51(8):1053-62.

39. Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today.* nov 2005;10(22):1551-8.

40. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol.* juill 2016;17(7):451-9.

41. Guijas C, Montenegro-Burke JR, Warth B, Spilker ME, Siuzdak G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat Biotechnol.* 5 avr 2018;36(4):316-20.

42. Xiang Z, Wang X, Cai X, Zeng S. Metabolomics Study on Quality Control and Discrimination of Three Curcuma Species based on Gas Chromatograph-Mass Spectrometry: Metabolomics Quality Control and Discrimination of Curcuma. *Phytochem Anal.* sept 2011;22(5):411-8.

43. Castillo Cáceres A. Evaluación agronómica de ajíes promisorios de la colección de germoplasma de capsicum del programa de hortalizas de la UNALM. [Lima]: Universidad Nacional Agraria La Molina; 2019.
44. Rojas R, Patel K, Ruiz C, Calderon R, Asencios E, Marceo M, et al. Ajies Nativos Peruanos. Caracterizacion agro-morfologica Quimico-Nutricional y Sensorial. 2016.
45. Fernie AR, Schauer N. Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet. janv 2009;25(1):39-48.
46. Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, et al. The genetics of plant metabolism. Nat Genet. juill 2006;38(7):842-9.
47. Tikunov Y. A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. PLANT Physiol. 1 nov 2005;139(3):1125-37.
48. Ritota M, Marini F, Sequi P, Valentini M. Metabolomic Characterization of Italian Sweet Pepper (*Capsicum annum* L.) by Means of HRMAS-NMR Spectroscopy and Multivariate Analysis. J Agric Food Chem. 8 sept 2010;58(17):9675-84.
49. Dunkler D, Sánchez-Cabo F, Heinze G. Statistical Analysis Principles for Omics Data. In: Mayer B, éditeur. Bioinformatics for Omics Data [Internet]. Totowa, NJ: Humana Press; 2011 [cité 18 mars 2019]. p. 113-31. Disponible sur: http://link.springer.com/10.1007/978-1-61779-027-0_5
50. Eggink, P.M., Wageningen University, Richard Visser, Arnaud Bovy, Chris Maliepaard, J.P.W. Haanstra. A taste of pepper: genetics, biochemistry and prediction of sweet pepper flavor [Internet]. s.n.; 2013 [cité 18 mars 2019]. Disponible sur: <http://library.wur.nl/WebQuery/wurpubs/443757>
51. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemom. mars 2002;16(3):119-28.

52. Wiklund S, Johansson E, Sjöström L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models. *Anal Chem.* janv 2008;80(1):115-22.
53. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom.* août 2006;20(8-10):341-51.
54. Sarstedt M, Ringle CM, Smith D, Reams R, Hair JF. Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *J Fam Bus Strategy.* mars 2014;5(1):105-15.
55. Eriksson L, Umetrics AB, éditeurs. Multi- and megavariate data analysis Basic principles and applications. third revised edition. Umeå: Umetrics; 2013. 503 p. (Multi- and megavariate data analysis).
56. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics.* 8 juin 2006;7(1):142.
57. Li B, Tang J, Yang Q, Li S, Cui X, Li Y, et al. NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* 3 juill 2017;45(W1):W162-70.
58. Kassambara A. Practical guide to principal component methods in R. Edition 1. United States}: CreateSpace Independent Publishing Platform; 2017. 155 p.
59. Projection-based clustering through self-organization and swarm intelligence. New York, NY: Springer Berlin Heidelberg; 2018.

60. Weimer BC, Slupsky C, éditeurs. *Metabolomics in food and nutrition*. Oxford; Philadelphia: Woodhead Publishing; 2013. 237 p. (Woodhead publishing series in food science, technology and nutrition).
61. Brown SM. *Essentials of medical genomics*. 2nd ed. Hoboken, N.J: Wiley-Blackwell; 2009. 439 p.
62. Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS® models. *J Chemom*. nov 2008;22(11-12):594-600.
63. Warnock DG, Peck CC. A roadmap for biomarker qualification. *Nat Biotechnol*. mai 2010;28(5):444-5.
64. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. juin 2006;27(8):861-74.
65. de Figueiredo M, Cordella CBY, Jouan-Rimbaud Bouveresse D, Archer X, Bégué J-M, Rutledge DN. A variable selection method for multiclass classification problems using two-class ROC analysis. *Chemom Intell Lab Syst*. juin 2018;177:35-46.
66. Calderón-Santiago M, Fernández-Peralbo MA, Priego-Capote F, Luque de Castro MD. MSCombine: a tool for merging untargeted metabolomic data from high-resolution mass spectrometry in the positive and negative ionization modes. *Metabolomics* [Internet]. mars 2016 [cité 19 mars 2019];12(3). Disponible sur: <http://link.springer.com/10.1007/s11306-016-0970-4>
67. Hervé MR, Nicolè F, Lê Cao K-A. *Multivariate Analysis of Multiple Datasets: a Practical Guide for Chemical Ecology*. *J Chem Ecol*. mars 2018;44(3):215-34.
68. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. Schneidman D, éditeur. *PLOS Comput Biol*. 3 nov 2017;13(11):e1005752.

69. Velickova E, Winkelhausen E, Kuzmanova S. Physical and sensory properties of ready to eat apple chips produced by osmo-convective drying. *J Food Sci Technol*. déc 2014;51(12):3691-701.
70. Boris S, Martin K, Matej P, Anna R, Jana M, Bohuslava T, et al. Application of multiple factor analysis for the descriptive sensory evaluation and instrumental measurements of bryndza cheese as affected by vacuum packaging. *Int J Food Prop*. janv 2018;21(1):1508-22.
71. Thomas E, Puget S, Valentin D, Songer P. Sensory Evaluation—Profiling and Preferences. In: *The Craft and Science of Coffee* [Internet]. Elsevier; 2017 [cité 5 août 2019]. p. 419-56. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780128035207000189>
72. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, et al. KNApSAcK Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research. *Plant Cell Physiol*. févr 2012;53(2):e1-e1.
73. Nakamura K, Shimura N, Otabe Y, Hirai-Morita A, Nakamura Y, Ono N, et al. KNApSAcK-3D: A Three-Dimensional Structure Database of Plant Metabolites. *Plant Cell Physiol*. févr 2013;54(2):e4-e4.
74. Ntie-Kang F, Telukunta KK, Döring K, Simoben CV, A. Moumbock AF, Malange YI, et al. NANPDB: A Resource for Natural Products from Northern African Sources. *J Nat Prod*. 28 juill 2017;80(7):2067-76.
75. Dreher K. Putting the Plant Metabolic Network Pathway Databases to Work: Going Offline to Gain New Capabilities. In: Sriram G, éditeur. *Plant Metabolism* [Internet]. Totowa, NJ: Humana Press; 2014 [cité 18 mars 2019]. p. 151-71. Disponible sur: http://link.springer.com/10.1007/978-1-62703-661-0_10

76. Gu J, Gui Y, Chen L, Yuan G, Lu H-Z, Xu X. Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. Cox D, éditeur. PLoS ONE. 25 avr 2013;8(4):e62839.
77. Zani CL, Carroll AR. Database for Rapid Dereplication of Known Natural Products Using Data from MS and Fast NMR Experiments. J Nat Prod. 23 juin 2017;80(6):1758-66.
78. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 23 nov 2012;41(D1):D456-63.
79. Schmelzer K, Fahy E, Subramaniam S, Dennis EA. The Lipid Maps Initiative in Lipidomics. In: Methods in Enzymology [Internet]. Elsevier; 2007 [cité 18 mars 2019]. p. 171-83. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0076687907320077>
80. Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, et al. LMSD: LIPID MAPS structure database. Nucleic Acids Res. 3 janv 2007;35(Database):D527-32.
81. Sud M, Fahy E, Cotter D, Dennis EA, Subramaniam S. LIPID MAPS-Nature Lipidomics Gateway: An Online Resource for Students and Educators Interested in Lipids. J Chem Educ. 10 janv 2012;89(2):291-2.
82. Buckingham J, éditeur. Dictionary of natural products. 1st ed. London ; New York: Chapman & Hall; 1994. 1 p.
83. Running W. Computer Software Reviews. Chapman and Hall Dictionary of Natural Products on CD-ROM. J Chem Inf Comput Sci. 1 nov 1993;33(6):934-5.
84. Samokhin A, Sotnezova K, Lashin V, Revelsky I. Evaluation of mass spectral library search algorithms implemented in commercial software: Evaluation of library search algorithms. J Mass Spectrom. juin 2015;50(6):820-5.

85. Journal of The American Society for Mass Spectrometry [Internet]. springer.com. [cité 18 mars 2019]. Disponible sur: <http://www.springer.com/chemistry/analytical+chemistry/journal/13361>
86. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 9 août 2016;34:828.
87. Meckelmann SW, Riegel DW, van Zonneveld MJ, Ríos L, Peña K, Ugas R, et al. Compositional Characterization of Native Peruvian Chili Peppers (*Capsicum* spp.). *J Agric Food Chem.* 13 mars 2013;61(10):2530-7.
88. MacBryde B. Consensus Document on the Biology of the *Capsicum annuum* Complex (Chili peppers, Hot peppers and Sweet peppers). 2006.
89. van Zonneveld M, Ramirez M, Williams DE, Petz M, Meckelmann S, Avila T, et al. Screening Genetic Resources of *Capsicum* Peppers in Their Primary Center of Diversity in Bolivia and Peru. *Vinutzer BA, éditeur. PLOS ONE.* 24 sept 2015;10(9):e0134663.
90. Eshbaugh WH. The taxonomy of the genus *Capsicum* (Solanaceae). *Phytologia.* 1980;47:153-66.
91. Libreros D, van Zonneveld M, Petz M, Meckelmann SW, Rios L, Amaya K, et al. Catálogo de ajíes (*Capsicum* spp.) peruanos promisorios conservados en el banco de semillas del INIA - Perú. *Bioversity International;* 2013.
92. International Plant Genetic Resources Institute, éditeur. Descriptors for *Capsicum* (*Capsicum* spp.) =: Descriptores para capsicum (*Capsicum* spp.). Rome: IPGRI; 1995. 51 p.
93. La Molina - Gaceta Molinera [Internet]. [cité 12 mai 2019]. Disponible sur: <http://www.lamolina.edu.pe/Gaceta/edicion2017/notas/nota079.htm>

94. Patel K, Ruiz C, Calderon R, Marcelo M, Rojas R. Characterisation of volatile profiles in 50 native Peruvian chili pepper using solid phase microextraction–gas chromatography mass spectrometry (SPME–GCMS). *Food Res Int.* nov 2016;89:471-5.
95. Spence C. Multisensory Flavor Perception. *Cell.* mars 2015;161(1):24-35.
96. Gibson M, Newsham P. Taste, Flavor and Aroma. In: *Food Science and the Culinary Arts* [Internet]. Elsevier; 2018 [cité 18 mars 2019]. p. 35-52. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780128118160000038>
97. Lim J, Fujimaru T, Linscott TD. The role of congruency in taste–odor interactions. *Food Qual Prefer.* juin 2014;34:5-13.
98. Choi N-E, Han JH. *How flavor works: the science of taste and aroma.* Chichester, West Sussex ; Hoboken, NJ, USA: John Wiley & Sons Inc; 2015. 223 p.
99. Noble AC. Taste-aroma interactions. *Trends Food Sci Technol.* déc 1996;7(12):439-44.
100. Zinngrebe Y. Learning from local knowledge in Peru—Ideas for more effective biodiversity conservation. *J Nat Conserv.* juill 2016;32:10-21.
101. Llerena R, Mendoza V, Ugás R, Canchari E. El punto de ají. *Investigaciones en Capsicum nativos* Números 1 y 2. 2012.
102. Prasad BCN, Kumar V, Gururaj HB, Parimalan R, Giridhar P, Ravishankar GA. Characterization of capsaicin synthase and identification of its gene (*csy1*) for pungency factor capsaicin in pepper (*Capsicum* sp.). *Proc Natl Acad Sci.* 5 sept 2006;103(36):13315.
103. ADEX P. Perú puede mejorar su posicionamiento como proveedor mundial de pimientos y ajíes [Internet]. ADEX. [cité 2 mars 2019]. Disponible sur: <http://www.adexperu.org.pe/notadeprensa/peru-puede-mejorar-su-posicionamiento-como-proveedor-mundial-de-pimientos-y-ajies/>

104. Jacobo Quispe Fredy. Caracterización Agro - Morfológica y Químico - Nutricional de Cinco Ajíes Promisorios del INIA - PERÚ. 2019° éd. Lima-Perú; 2019. 125 p. (Instituto Nacional de Innovación Agraria-INIA; vol. 1).
105. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. juin 2015;12(6):523-6.
106. Lai Z, Tsugawa H, Wohlgemuth G, Mehta S, Mueller M, Zheng Y, et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat Methods*. janv 2018;15(1):53-6.
107. Olivon F, Elie N, Grelier G, Roussi F, Litaudon M, Touboul D. MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm. *Anal Chem*. 4 déc 2018;90(23):13900-8.
108. Moreno E, Fita A, González-Mas MC, Rodríguez-Burruezo A. HS-SPME study of the volatile fraction of Capsicum accessions and hybrids in different parts of the fruit. *Sci Hortic*. févr 2012;135:87-97.
109. Thissen U, Coulier L, Overkamp KM, Jetten J, van der Werff BJC, van de Ven T, et al. A proper metabolomics strategy supports efficient food quality improvement: A case study on tomato sensory properties. *Food Qual Prefer*. sept 2011;22(6):499-506.
110. Garruti D dos S, Pinto N de OF, Alves VCC, Penha MFA da, Tobaruela E de C, Araújo ÍM da S. Volatile profile and sensory quality of new varieties of Capsicum chinense pepper. *Ciênc E Tecnol Aliment*. févr 2013;33:102-8.
111. Mickiewicz B, Tam P, Jenne CN, Leger C, Wong J, Winston BW, et al. Integration of metabolic and inflammatory mediator profiles as a potential prognostic approach for septic shock in the intensive care unit. *Crit Care*. 2015;19(1):11.

112. Wheelock AM, Wheelock CE. Trials and tribulations of 'omics data analysis: assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. *Mol Biosyst.* 2013;9(11):2589.
113. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, et al. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol Biosyst.* 2015;11(1):13-9.
114. Tugizimana F, Steenkamp P, Piater L, Dubery I. A Conversation on Data Mining Strategies in LC-MS Untargeted Metabolomics: Pre-Processing and Pre-Treatment Steps. *Metabolites.* 3 nov 2016;6(4):40.
115. Montgomery DC, Runger GC. Applied statistics and probability for engineers. Sixth edition. Hoboken, NJ: John Wiley and Sons, Inc; 2014. 811 p.
116. Lawal B. Applied Statistical Methods in Agriculture, Health and Life Sciences [Internet]. Cham: Springer International Publishing; 2014 [cité 8 sept 2019]. Disponible sur: <http://link.springer.com/10.1007/978-3-319-05555-8>
117. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. *Chemom Intell Lab Syst.* juill 2005;78(1-2):103-12.
118. Karpievitch YV, Nikolic SB, Wilson R, Sharman JE, Edwards LM. Metabolomics Data Normalization with EigenMS. Dalby AR, éditeur. *PLoS ONE.* 30 déc 2014;9(12):e116221.
119. Karpievitch YV, Taverner T, Adkins JN, Callister SJ, Anderson GA, Smith RD, et al. Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. *Bioinformatics.* 1 oct 2009;25(19):2573-80.
120. Shah JS, Rai SN, DeFilippis AP, Hill BG, Bhatnagar A, Brock GN. Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies. *BMC Bioinformatics* [Internet]. déc 2017

- [cité 25 juill 2019];18(1). Disponible sur:
<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1547-6>
121. The Human Serum Metabolome (HUSERMET) Consortium, Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc.* juill 2011;6(7):1060-83.
122. Sandasi M, Vermaak I, Chen W, Viljoen A. The Application of Vibrational Spectroscopy Techniques in the Qualitative Assessment of Material Traded as Ginseng. *Molecules.* 12 avr 2016;21(4):472.
123. Marti G, Joulia P, Amiel A, Fabre B, David B, Fabre N, et al. Comparison of the Phytochemical Composition of *Serenoa repens* Extracts by a Multiplexed Metabolomic Approach. *Molecules.* 13 juin 2019;24(12):2208.
124. Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P. A Sparse PLS for Variable Selection when Integrating Omics Data. *Stat Appl Genet Mol Biol* [Internet]. 18 janv 2008 [cité 24 oct 2019];7(1). Disponible sur:
<https://www.degruyter.com/view/j/sagmb.2008.7.1/sagmb.2008.7.1.1390/sagmb.2008.7.1.1390.xml>
125. Chung D, Keles S. Sparse Partial Least Squares Classification for High Dimensional Data. *Stat Appl Genet Mol Biol* [Internet]. 3 janv 2010 [cité 24 oct 2019];9(1). Disponible sur:
<https://www.degruyter.com/view/j/sagmb.2010.9.1/sagmb.2010.9.1.1492/sagmb.2010.9.1.1492.xml>
126. Tenenbaum G, Filho E. Chapter 3 - Measurement Considerations in Performance Psychology. In: Raab M, Lobinger B, Hoffmann S, Pizzera A, Laborde S, éditeurs. *Performance Psychology* [Internet]. San Diego: Academic Press; 2016. p. 31-44.

Disponible

sur:

<http://www.sciencedirect.com/science/article/pii/B978012803377700003X>

127. Calvini R, Ulrici A, Amigo JM. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemom Intell Lab Syst.* août 2015;146:503-11.

128. Barbin DF, ElMasry G, Sun D-W, Allen P. Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Anal Chim Acta.* mars 2012;719:30-42.

129. González I, Cao K-AL, Davis MJ, Déjean S. Visualising associations between paired 'omics' data sets. *BioData Min [Internet].* déc 2012 [cité 30 oct 2019];5(1). Disponible sur: <https://biodatamining.biomedcentral.com/articles/10.1186/1756-0381-5-19>

130. Bicho NC, Leitão AE, Ramalho JC, de Alvarenga NB, Lidon FC. Impact of Roasting Time on the Sensory Profile of Arabica and Robusta Coffee. *Ecol Food Nutr.* mars 2013;52(2):163-77.

131. Rocha IF de O, Bolini HMA. Passion fruit juice with different sweeteners: sensory profile by descriptive analysis and acceptance. *Food Sci Nutr.* mars 2015;3(2):129-39.

132. Kuti T, Hegyi A, Kemény S. Analysis of sensory data of different food products by ANOVA. *Chemom Intell Lab Syst.* juill 2004;72(2):253-7.

133. Brown AM. A new software for carrying out one-way ANOVA post hoc tests. *Comput Methods Programs Biomed.* juill 2005;79(1):89-95.

134. Lahne J. Sensory science, the food industry, and the objectification of taste. *Anthropol Food [Internet].* 28 janv 2016 [cité 4 déc 2019];(10). Disponible sur: <http://journals.openedition.org/aof/7956>

135. Heymann H, Machado B, Torri L, Robinson AL. HOW MANY JUDGES SHOULD ONE USE FOR SENSORY DESCRIPTIVE ANALYSIS?: HOW MANY JUDGES FOR SENSORY DESCRIPTIVE ANALYSIS? *J Sens Stud.* avr 2012;27(2):111-22.
136. Kirti Patel, Rosa Calderon, Edgard Asencios, Dioliza Vilchez, Mavel Marcelo, Rosario Rojas. Agro-morphological Characteristics and Sensory Evaluation of Native Peruvian Chili Peppers. *J Agric Sci Technol B* [Internet]. 28 sept 2016 [cité 4 déc 2019];6(3). Disponible sur: <http://www.davidpublisher.org/index.php/Home/Article/index?id=30886.html>
137. Tacq J. Multivariate Normal Distribution. In: *International Encyclopedia of Education* [Internet]. Elsevier; 2010 [cité 6 nov 2019]. p. 332-8. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780080448947013518>
138. Ballabio D, Todeschini R. Multivariate Classification for Qualitative Analysis. In: *Infrared Spectroscopy for Food Quality Analysis and Control* [Internet]. Elsevier; 2009 [cité 2 déc 2019]. p. 83-104. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780123741363000043>
139. Homayoun SB, Shrikant IB, Kazem M, Hemmat M, Reza M. Compared application of the new OPLS-DA statistical model versus partial least squares regression to manage large numbers of variables in an injury case-control study. *Sci Res Essays.* 19 sept 2011;6(20):4369-77.
140. Lindinger C, Labbe D, Pollien P, Rytz A, Juillerat MA, Yeretziyan C, et al. When Machine Tastes Coffee: Instrumental Approach To Predict the Sensory Profile of Espresso Coffee. *Anal Chem.* mars 2008;80(5):1574-81.
141. Sukha DA, Butler DR, Umaharan P, Boulton E. The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. *Eur Food Res Technol.* janv 2008;226(3):405-13.

142. Yang L-N, Pu J-C, Liu L-X, Wang G-W, Zhou X-Y, Zhang Y-Q, et al. Integrated Metabolomics and Proteomics Analysis Revealed Second Messenger System Disturbance in Hippocampus of Chronic Social Defeat Stress Rat. *Front Neurosci* [Internet]. 22 mars 2019 [cité 4 déc 2019];13. Disponible sur: <https://www.frontiersin.org/article/10.3389/fnins.2019.00247/full>
143. Baumann K, Carnicer M, Dragosits M, Graf AB, Stadlmann J, Jouhten P, et al. A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst Biol*. 2010;4(1):141.
144. Bourn D, Prescott J. A Comparison of the Nutritional Value, Sensory Qualities, and Food Safety of Organically and Conventionally Produced Foods. *Crit Rev Food Sci Nutr*. janv 2002;42(1):1-34.
145. Huber M, Rembiałkowska E, Średnicka D, Bügel S, van de Vijver LPL. Organic food and impact on human health: Assessing the status quo and prospects of research. *NJAS - Wagening J Life Sci*. déc 2011;58(3-4):103-9.
146. Williams CM. Nutritional quality of organic food: shades of grey or shades of green? *Proc Nutr Soc*. févr 2002;61(1):19-24.
147. Shennan C, Krupnik TJ, Baird G, Cohen H, Forbush K, Lovell RJ, et al. Organic and Conventional Agriculture: A Useful Framing? *Annu Rev Environ Resour*. 17 oct 2017;42(1):317-46.
148. Rodríguez-García C, Sánchez-Quesada C, Gaforio JJ. Dietary Flavonoids as Cancer Chemopreventive Agents: An Updated Review of Human Studies. *Antioxidants*. 18 mai 2019;8(5):137.
149. Huang M, Lu J-J, Huang M-Q, Bao J-L, Chen X-P, Wang Y-T. Terpenoids: natural products for cancer therapy. *Expert Opin Investig Drugs*. déc 2012;21(12):1801-18.

150. Siasos G, Tousoulis D, Tsigkou V, Kokkou E, Oikonomou E, Vavuranakis M, et al. Flavonoids in Atherosclerosis: An Overview of Their Mechanisms of Action. *Curr Med Chem*. 1 mai 2013;20(21):2641-60.
151. Shaito A, Thuan DTB, Phu HT, Nguyen THD, Hasan H, Halabi S, et al. Herbal Medicine for Cardiovascular Diseases: Efficacy, Mechanisms, and Safety. *Front Pharmacol* [Internet]. 7 avr 2020 [cité 9 sept 2020];11. Disponible sur: <https://www.frontiersin.org/article/10.3389/fphar.2020.00422/full>
152. Lalani S, Poh CL. Flavonoids as Antiviral Agents for Enterovirus A71 (EV-A71). *Viruses*. 6 févr 2020;12(2):184.
153. Isah T. Anticancer alkaloids from trees: Development into drugs. *Pharmacogn Rev*. 2016;10(20):90.
154. Traka M, Mithen R. Glucosinolates, isothiocyanates and human health. *Phytochem Rev*. janv 2009;8(1):269-82.
155. Rembiałkowska E. Organic Food: Effect on Nutrient Composition. In: *Encyclopedia of Food and Health* [Internet]. Elsevier; 2016 [cité 8 sept 2020]. p. 171-7. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780123849472005055>
156. Duminil J, Di Michele M. Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosyst - Int J Deal Asp Plant Biol*. nov 2009;143(3):528-42.
157. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* [Internet]. juin 2018 [cité 5 déc 2019];14(6). Disponible sur: <http://link.springer.com/10.1007/s11306-018-1367-3>

158. Gika HG, Theodoridis GA, Wingate JE, Wilson ID. Within-Day Reproducibility of an HPLC–MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J Proteome Res.* août 2007;6(8):3291-303.
159. Livera AMD, Sysi-Aho M, Jacob L, Gagnon-Bartsch JA, Castillo S, Simpson JA, et al. Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal Chem.* 7 avr 2015;87(7):3606-15.
160. Li X-N, Tong S-Q, Cheng D-P, Li Q-Y, Yan J-Z. Coumarins from *Edgeworthia chrysantha*. *Molecules.* 13 févr 2014;19(2):2042-8.
161. Brown EG, Davies D. Extraction, pre-high-performance liquid chromatographic purification, and high-performance liquid chromatographic analysis of plant nucleotides. *Anal Biochem.* déc 1989;183(2):312-9.
162. Bae H, Jayaprakasha GK, Jifon J, Patil BS. Extraction efficiency and validation of an HPLC method for flavonoid analysis in peppers. *Food Chem.* févr 2012;130(3):751-8.
163. Fernández-Trujillo JP. Extraction of sweet and hot pepper and paprika oleoresin I. Overview, composition, process, innovations, and applications. *Grasas Aceites* [Internet]. 30 sept 2007 [cité 6 déc 2019];58(3). Disponible sur: <http://grasasyaceites.revistas.csic.es/index.php/grasasyaceites/article/view/180/180>
164. Petruczynik A. Analysis of alkaloids from different chemical groups by different liquid chromatography methods. *Open Chem* [Internet]. 1 janv 2012 [cité 6 déc 2019];10(3). Disponible sur: <http://www.degruyter.com/view/j/chem.2012.10.issue-3/s11532-012-0037-y/s11532-012-0037-y.xml>
165. Marti G, Erb M, Rudaz S, Turlings T, Wolfender J-L. Search for Low-Molecular-Weight Biomarkers in Plant Tissues and Seeds Using Metabolomics: Tools, Strategies, and Applications. In: Agrawal GK, Rakwal R, éditeurs. *Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield* [Internet]. Dordrecht:

Springer Netherlands; 2012 [cité 12 déc 2019]. p. 305-41. Disponible sur: http://link.springer.com/10.1007/978-94-007-4749-4_16

166. Chervin J, Perio P, Martins-Froment N, Pharkeovilay C, Reybier K, Nepveu F, et al. Dereplication of natural products from complex extracts by regression analysis and molecular networking: case study of redox-active compounds from *Viola alba* subsp. *dehnhardtii*. *Metabolomics* [Internet]. août 2017 [cité 7 déc 2019];13(8). Disponible sur: <http://link.springer.com/10.1007/s11306-017-1227-6>

167. Villa-Ruano N, Velásquez-Valle R, Zepeda-Vallejo LG, Pérez-Hernández N, Velázquez-Ponce M, Arcos-Adame VM, et al. ¹H NMR-based metabolomic profiling for identification of metabolites in *Capsicum annuum* cv. mirasol infected by beet mild curly top virus (BMCTV). *Food Res Int.* avr 2018;106:870-7.

168. Macel M, Visschers IGS, Peters JL, Kappers IF, de Vos RCH, van Dam NM. Metabolomics of Thrips Resistance in Pepper (*Capsicum* spp.) Reveals Monomer and Dimer Acyclic Diterpene Glycosides as Potential Chemical Defenses. *J Chem Ecol.* juin 2019;45(5-6):490-501.

169. Florentino-Ramos E, Villa-Ruano N, Hidalgo-Martínez D, Ramírez-Meraz M, Méndez-Aguilar R, Velásquez-Valle R, et al. ¹H NMR-based fingerprinting of eleven Mexican *Capsicum annuum* cultivars. *Food Res Int.* juill 2019;121:12-9.

170. Cervantes-Hernández F, Alcalá-González P, Martínez O, Ordaz-Ortiz JJ. Placenta, Pericarp, and Seeds of Tabasco Chili Pepper Fruits Show a Contrasting Diversity of Bioactive Metabolites. *Metabolites.* 28 sept 2019;9(10):206.

171. Knudsen JT, Eriksson R, Gershenzon J, Ståhl B. Diversity and Distribution of Floral Scent. *Bot Rev.* mars 2006;72(1):1-120.

172. Pinto NOF, Castro Alves VC, Penha MFA, Gomes BL, Reifschneider FJB, Garruti DS. Volatile Profile of New Improved Brazilian *Capsicum chinense* Peppers. In: *Flavour*

Science [Internet]. Elsevier; 2014 [cité 11 déc 2019]. p. 643-7. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780123985491001185>

173. Kollmannsberger H, Rodríguez-Burruezo A, Nitz S, Nuez F. Volatile and capsaicinoid composition of ají (*Capsicum baccatum*) and rocoto (*Capsicum pubescens*), two Andean species of chile peppers. *J Sci Food Agric*. juill 2011;91(9):1598-611.

174. Rodríguez-Burruezo A, Kollmannsberger H, González-Mas MC, Nitz S, Fernando N. HS-SPME Comparative Analysis of Genotypic Diversity in the Volatile Fraction and Aroma-Contributing Compounds of *Capsicum* Fruits from the *annuum* – *chinense* – *frutescens* Complex. *J Agric Food Chem*. 14 avr 2010;58(7):4388-400.

175. Luning PA, de Rijk T, Wichers HJ, Roozen JP. Gas Chromatography, Mass Spectrometry, and Sniffing Port Analyses of Volatile Compounds of Fresh Bell Peppers (*Capsicum annuum*) at Different Ripening Stages. *J Agric Food Chem*. avr 1994;42(4):977-83.

176. Forero MD, Quijano CE, Pino JA. Volatile compounds of chile pepper (*Capsicum annuum* L. var. *glabriusculum*) at two ripening stages. *Flavour Fragr J*. janv 2009;24(1):25-30.

177. de Torres C, Díaz-Maroto MC, Hermosín-Gutiérrez I, Pérez-Coello MS. Effect of freeze-drying and oven-drying on volatiles and phenolics composition of grape skin. *Anal Chim Acta*. févr 2010;660(1-2):177-82.

178. Ma C, Tan C, Li W, Chen L, Wang Y, Chen X. Identification of the Different Aroma Compounds between Conventional and Freeze Dried Wuyi Rock Tea (Dangui) using Headspace Solid Phase Microextraction. *Food Sci Technol Res*. 2013;19(5):805-11.

179. Zhang L, Qiao Y, Wang C, Liao L, Liu L, Shi D, et al. Effects of Freeze Vacuum Drying Combined with Hot Air Drying on the Sensory Quality, Active Components,

Moisture Mobility, Odors, and Microstructure of Kiwifruits. *J Food Qual.* 7 mars 2019;2019:1-11.

180. Youn-Soon Kim, Hye-Young Seo. Comparison of Volatile Organic Components in Fresh and Freeze Dried Garlic. *J Korean Soc Food Sci Nutr.* 1 juill 2005;34(6):885-91.

181. Díaz-Maroto MC, Sánchez Palomo E, Castro L, González Viñas M, Pérez-Coello MS. Changes produced in the aroma compounds and structural integrity of basil (*Ocimum basilicum* L) during drying: Effect of drying on flavour quality of basil. *J Sci Food Agric.* déc 2004;84(15):2070-6.

182. Oruna-Concha MJ, Lopez-Hernandez J, Simal-Lozano JA, Simal-Gandara J, Gonzalez-Castro MJ, Garcia C d. I. C. Determination of Volatile Components in Fresh, Frozen, and Freeze-Dried Padron-Type Peppers by Gas Chromatography--Mass Spectrometry Using Dynamic Headspace Sampling and Microwave Desorption. *J Chromatogr Sci.* 1 déc 1998;36(12):583-8.

183. Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, et al. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* 28 févr 2005;579(6):1332-7.

184. Jang YK, Jung ES, Lee H-A, Choi D, Lee CH. Metabolomic Characterization of Hot Pepper (*Capsicum annuum* "CM334") during Fruit Development. *J Agric Food Chem.* 4 nov 2015;63(43):9452-60.

185. Debnath M, De B, Das S. GC-MS-Based Profiling of Non-polar Metabolites and Chemometric Study of Fruits of *Capsicum* Species and Landraces at Different Stages of Ripening. *J Herbs Spices Med Plants.* 19 nov 2019;1-22.

186. Hirokawa Y, Nheu T, Grimm K, Mautner V, Maeda S, Yoshida M, et al. Sichuan pepper extracts block the PAK1/Cyclin D1 pathway and the growth of NF1-deficient cancer xenograft in mice. *Cancer Biol Ther.* mars 2006;5(3):305-9.

187. Amarowicz R. Antioxidant activity of peppers: Editorial. *Eur J Lipid Sci Technol.* mars 2014;116(3):237-9.
188. Russo VM, éditeur. *Peppers: botany, production and uses.* Wallingford, Oxfordshire, UK ; Cambridge, MA: CABI; 2012. 280 p.
189. Bhatia A, Sarma SJ, Lei Z, Sumner LW. UHPLC-QTOF-MS/MS-SPE-NMR: A Solution to the Metabolomics Grand Challenge of Higher-Throughput, Confident Metabolite Identifications. In: Gowda GAN, Raftery D, éditeurs. *NMR-Based Metabolomics* [Internet]. New York, NY: Springer New York; 2019 [cité 9 sept 2020]. p. 113-33. (Methods in Molecular Biology; vol. 2037). Disponible sur: http://link.springer.com/10.1007/978-1-4939-9690-2_7
190. Vásquez-Ocmín PG, Gadea A, Cojean S, Marti G, Pomel S, Van Baelen A-C, et al. Metabolomic approach of the antiprotozoal activity of medicinal Piper species used in Peruvian Amazon. *J Ethnopharmacol.* janv 2021;264:113262.
191. Oliveri P, Simonetti R. Chemometrics for Food Authenticity Applications. In: *Advances in Food Authenticity Testing* [Internet]. Elsevier; 2016 [cité 12 déc 2019]. p. 701-28. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780081002209000254>
192. Granato D, Putnik P, Kovačević DB, Santos JS, Calado V, Rocha RS, et al. Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing: Trends in chemometrics.... *Compr Rev Food Sci Food Saf.* mai 2018;17(3):663-77.
193. Novotná H, Kmiecik O, Gałązka M, Krtková V, Hurajová A, Schulzová V, et al. Metabolomic fingerprinting employing DART-TOFMS for authentication of tomatoes and peppers from organic and conventional farming. *Food Addit Contam Part A.* sept 2012;29(9):1335-46.

194. Bakeev KA, éditeur. Process analytical technology: spectroscopic tools and implemented strategies for the chemical and pharmaceutical industries. 2nd ed. Chichester, West Sussex: Wiley; 2010. 557 p.
195. Barker M, Rayens W. Partial least squares for discrimination. *J Chemom.* 24 mars 2003;17(3):166-73.
196. Abdul Rashid N, Siti Esah Che Hussain W, Razak Ahmad A, Norazami Abdullah F. Performance of Classification Analysis: A Comparative Study between PLS-DA and Integrating PCA+LDA. *Math Stat.* sept 2019;7(4A):24-8.
197. Marcello Manfredi ER. Biomarkers Discovery through Multivariate Statistical Methods: A Review of Recently Developed Methods and Applications in Proteomics. *J Proteomics Bioinform* [Internet]. 2013 [cité 13 déc 2019];s3. Disponible sur: <http://omicsonline.org/open-access/biomarkers-discovery-through-multivariate-statistical-methods-jpb.S3-003.php?aid=33716>
198. Cubero-Leon E, De Rudder O, Maquet A. Metabolomics for organic food authentication: Results from a long-term field study in carrots. *Food Chem.* janv 2018;239:760-70.
199. Navratilova K, Hrbek V, Kratky F, Hurkova K, Tomaniova M, Pulkrabova J, et al. Green tea: Authentication of geographic origin based on UHPLC-HRMS fingerprints. *J Food Compos Anal.* mai 2019;78:121-8.
200. Varrà MO, Ghidini S, Zanardi E, Badiani A, Ianieri A. Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. *Ital J Food Saf* [Internet]. 22 mars 2019 [cité 14 déc 2019];8(1). Disponible sur: <https://www.pagepressjournals.org/index.php/ijfs/article/view/7872>

201. de Wijk RA, Prinz JF, Engelen L, Weenen H. The role of α -amylase in the perception of oral texture and flavour in custards. *Physiol Behav.* oct 2004;83(1):81-91.
202. Sales PM, Souza PM, Simeoni LA, Magalhães PO, Silveira D. α -Amylase Inhibitors: A Review of Raw Material and Isolated Compounds from Plant Source. *J Pharm Pharm Sci.* 10 févr 2012;15(1):141.
203. Stalikas CD. Extraction, separation, and detection methods for phenolic acids and flavonoids. *J Sep Sci.* déc 2007;30(18):3268-95.
204. Manach C, Scalbert A, Morand C, Rémésy C, Jiménez L. Polyphenols: food sources and bioavailability. *Am J Clin Nutr.* 1 mai 2004;79(5):727-47.
205. Bassoli A, Borgonovo G, Busnelli G. Alkaloids and the Bitter Taste. In: Fattorusso E, Tagliatela-Scafati O, éditeurs. *Modern Alkaloids* [Internet]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2007 [cité 16 déc 2019]. p. 53-72. Disponible sur: <http://doi.wiley.com/10.1002/9783527621071.ch3>
206. Chéron J-B, Marchal A, Fiorucci S. Natural Sweeteners. In: *Encyclopedia of Food Chemistry* [Internet]. Elsevier; 2019 [cité 16 déc 2019]. p. 189-95. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B978008100596521620X>
207. Musialik M, Kuzmich R, Pawłowski TS, Litwinienko G. Acidity of Hydroxyl Groups: An Overlooked Influence on Antiradical Properties of Flavonoids. *J Org Chem.* 3 avr 2009;74(7):2699-709.
208. Zengin G, Ferrante C, Gnapi DE, Sinan KI, Orlando G, Recinella L, et al. Comprehensive approaches on the chemical constituents and pharmacological properties of flowers and leaves of American basil (*Ocimum americanum* L). *Food Res Int.* nov 2019;125:108610.

209. Yin Q, Hung S-C, Rathmell WK, Shen L, Wang L, Lin W, et al. Integrative radiomics expression predicts molecular subtypes of primary clear cell renal cell carcinoma. *Clin Radiol.* sept 2018;73(9):782-91.

13. ANEXO

13.1.ANEXO 1 (Perfil de metabolitos No volátiles)

Tabla 60. Áreas relativas de metabolitos no volátiles de la accesión Ayuclo

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	t _R (min).	promedio de áreas
Neg	Citric acid	88	All-public	191.01332	0.747	77650144
POS	Capsaicin	7.4281	DNP-Capsicum	306.20718	7.27	45100522
POS	Dihydrocapsaicin	0.96	GNPS	308.22189	7.725	42252761
POS	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	7.1793	DNP-Capsicum	146.07973	0.706	36640393
Neg	(S)-MALATE	88	All-public	133.00992	0.747	22602758
Neg	L-Lactic acid	6.0312	Msfinder	225.05371	0.67	21053049
POS	Capsaicin; Nonanoyl analogue	6.8459	DNP-Capsicum	294.20636	7.173	15346944.5
POS	Phenylalanine; (R)-form	7.3057	DNP-Solanaceae	166.08464	1.46	12970432.75
POS	Isoleucine; (2R,3R)-form	7.7789	DNP-Solanaceae	132.10057	1.012	9731968.75
POS	Ethyl-1,4-benzoquinone	7.6235	DNP-Solanaceae	137.05931	7.27	8930075.25
POS	2,5-Dimethyl-1,4-benzoquinone	6.8901	DNP-Capsicum	137.0594	7.725	8847008.25
POS	Acetamide; N-(13-Methyltetradecyl)	7.0194	DNP-Solanaceae	270.27728	10.6	8340236.5
POS	Tryptophan	88	All-public	205.09535	2.532	7291634.5
POS	L-Asparagine	85.7	All-public	133.05951	0.731	5750798.25
POS	L-(+)-Arginine	96.3	All-public	175.11726	0.669	5080518.5
POS	Phosphocholine	0.91	GNPS	184.03532	0.669	4954670.75
Neg	6-O-a-D-Glucopyranuronosyl-D-glucose; a-Pyranose-form	7.708	DNP-Solanaceae	355.09122	3.247	4001426.25
Neg	5-Hydroxy-2-(4-hydroxyphenyl) benzofuran	6.6683	DNP-Solanaceae	225.05373	1.043	3895903.75
POS	6-Dodecyl-3-hydroxy-2-methylpyridine; 7'-Oxo, 12'-hydroxy	6.2256	DNP-Solanaceae	615.43665	7.725	3632956
POS	Capsaicin; (Z)-isomer	6.2572	DNP-Capsicum	611.40723	7.27	3602221.75

Tabla 61. Continuación tabla 60...

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Db	Mz	t _R (min)	promedio de áreas
Neg	Dehydroascorbic acid	8.8165	Msfinder	173.00336	1.221	3554634.5
POS	NCGC00385770-01!2-(3,4-dihydroxyphenyl)-5-hydroxy-7-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[(3,4,5-trihydroxyoxan-2-yl)oxymethyl]oxan-2-yl]oxychromen-4-one	87.8	All-public	581.14618	3.909	3447861.25
Neg	quercitrin	0.91	GNPS	447.07776	4.226	3429692.75
Neg	3-(3,4-Dihydroxyphenyl)-2-propen-1-ol; 3'-Me ether, 1-cinnamoyl	7.5606	DNP-Solanaceae	309.10843	3.182	3372074.75
POS	L-Phenylalanine	0.93	GNPS	120.07944	1.46	2917853.5
POS	3,20-Epoxy-18-nor-15-beyerene-3,6-diol; (3aOH,6a)-form	7.9478	DNP-Capsicum	322.23776	8.252	2893511
Neg	3,4,6-Phenanthrenetriol; 3-O-β-D-Glucopyranoside	7.3007	DNP-Capsicum	387.10095	0.696	2883012
POS	Capsaicin; 9-Methyl-7E-decenoyl analogue	7.8176	DNP-Capsicum	320.22195	7.804	2796893.25
POS	9-Oxo-10E,12Z-octadecadienoic	0.9	GNPS	295.11127	0.731	2753110
Neg	Hypochoeroside D;(+)Hypochoeroside D	7.6496	Msfinder	567.18835	3.843	2601583.5
POS	Capsaicin; Octanoyl analogue	7.7434	DNP-Capsicum	280.18811	6.697	2459493.5
POS	Glucuronic acid; 6-Aminoheyl glycoside	7.3598	DNP-Solanaceae	294.15213	1.005	2348799.5
POS	Quercetin-3-O-rhamnoside	92.6	All-public	449.10471	4.214	2136114
Neg	3,3',4',7-Tetrahydroxyflavanone; 3-Ac	7.2802	DNP-Solanaceae	329.07635	2.452	2024448.75
POS	Capsaicin; Decanoyl analogue	7.4288	DNP-Capsicum	330.20401	7.725	1938651.75
POS	3-Hydroxynonanoic acid; Hydrazide	7.6397	DNP-Capsicum	189.15808	0.649	1841176.25
Neg	4-Amino-4-deoxyglucose; 2,3:5,6-Di-O-isopropylidene, di-Me acetal	7.3346	DNP-Solanaceae	350.18527	7.288	1810097.5
POS	Uridine	0.84	GNPS	171.01511	0.675	1777264.5
POS	Deacetylfaucettine; 8-Ketone, 5-Ac	6.3446	DNP-Solanaceae	633.38702	7.27	1684081.25
POS	Paniculine	6.4522	DNP-Solanaceae	637.41846	7.725	1682717.5

Tabla 62. Áreas relativas de metabolitos no volátiles de la accesión Charapita

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	tr (min)	promedio de áreas	%RSD
POS	Capsaicin	7.9242	Msfinder	306.20627	7.279	121517811.1	19.17
POS	Dihydrocapsaicin	84.4	all-public	308.22208	7.719	60416365.54	13.85
Neg	(S)-MALATE	88	all-public	133.00992	0.743	46144158.55	7.76
Neg	Citric acid	87.9	all-public	191.01323	0.896	31849168.98	14.08
POS	2,3-Dihydro-5-benzofuranol	7.2367	DNP-Solanaceae	137.05936	7.272	24207148.65	20.53
POS	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	7.3737	DNP-Capsicum	146.07964	0.712	20032562.14	12.40
POS	2,5-Dimethyl-1,4-benzoquinone	6.991	DNP-Capsicum	137.05946	7.719	12418813.54	11.42
POS	2,4,5-Trihydroxy-2-(hydroxymethyl)pentanoic acid	7.455	DNP-Solanaceae	219.02444	0.693	6836358.539	18.39
POS	2-Hydroxybenzylamine; N-Ac	7.7224	DNP-Capsicum	166.08461	1.464	6286796.893	12.68
Neg	2,3-Dihydro-5,7-dihydroxy-2-methyl-4H-1-benzopyran-4-one	6.388	DNP-Solanaceae	215.02542	0.685	5333789.137	6.30
POS	Capsaicin; Nonanoyl analogue	7.8399	DNP-Capsicum	294.20633	7.197	4443091.039	5.28
POS	Lankamycin; 3-O-Deglycosyl	7.0519	DNP-Solanaceae	633.38708	7.272	4337711.854	16.78
POS	Aminocaproic acid	87.5	all-public	132.10049	1.019	4130342.604	14.61
POS	L-Asparagine	84.7	all-public	133.05943	0.736	3859263.947	9.15
POS	Tryptophan	88	all-public	205.09535	2.518	3719326.104	17.62
Neg	3,3',4',7-Tetrahydroxyflavanone; 3-Ac	7.4372	DNP-Solanaceae	329.07642	2.451	3424363.29	13.78
Neg	Nimbin; 21,23-Dihydro, 23-oxo, 21 β -hydroxy	6.6049	DNP-Solanaceae	571.2229	4.632	3250724.726	8.54
POS	N-Cinnamoylputrescine ; 3",4"-Dihydroxy	7.7967	DNP-Solanaceae	251.13683	2.036	2655347.007	11.92
Neg	3-Phenyl-2-propen-1-ol; O-[α -L-Rhamnopyranosyl-(1 \rightarrow 6)- β -D-glucopyranoside]	7.5662	DNP-Capsicum	441.1821	4.068	2581680.214	10.04

Tabla 63. Continuación tabla 62...

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	tr (min).	promedio de áreas
Neg	8,9,10-Trihydroxy-1-oxo-2,11(13)-germacradien-12,6-olide; 8,9-Bis-O-(2-methylpropenoyl)	7.3535	DNP-Solanaceae	431.17722	3.445	2552701.089
POS	Lycoclavine	6.3613	DNP-Solanaceae	330.20419	7.719	2397082.944
POS	Capsaicin; Decanoyl analogue	6.3463	DNP-Capsicum	637.41943	7.715	2363702.143
Neg	quercitrin	0.91	GNPS	447.0777	4.219	2065937.044
POS	8,14-Dihydroxy-19-nor-4(18),15-isopimaradien-7-one; (8?,14?)-form	7.8187	DNP-Capsicum	322.2374	8.244	2052117.049
POS	2,4-Pentanediol; 4-O-(4-Hydroxy-E-cinnamoyl)	7.6533	DNP-Capsicum	251.13672	1.446	1999453.089
POS	Phosphocholine	0.93	GNPS	184.03526	0.674	1923632.735
POS	Mescaline; N-Ac	7.6175	DNP-Capsicum	254.13628	2.368	1904953.074
Neg	3-(3,4-Dihydroxyphenyl)-2-propen-1-ol; 3'-Me ether, 1-cinnamoyl	7.4945	DNP-Capsicum	309.10834	3.192	1890584.838
Neg	6-O-a-D-Glucopyranuronosyl-D-glucose; a-Pyranose-form	7.6021	DNP-Capsicum	355.09116	3.247	1849964.849
POS	N-Cinnamoylputrescine ; 3"-Methoxy, 4"-hydroxy	7.3997	DNP-Capsicum	265.15247	2.611	1841944.922
POS	4-Hydroxynonanoic acid; Hydrazide	7.6297	DNP-Solanaceae	189.1581	0.654	1507038.931
POS	Isorhamnetin 3-glycosides; 3-O-[?-D-Xylopyranosyl-(1→3)-a-L-rhamnopyranosyl-(1→6)-β-D-galactopyranoside]	6.8269	DNP-Solanaceae	757.21289	4.177	1505507.435
POS	[4-(Aminomethyl)phenyl]methanamine	0.72	GNPS	120.07948	1.464	1465982.893
Neg	Homotyramine; Me ether, N-(3-hydroxy-4-methoxy-E-cinnamoyl)	6.2035	DNP-Capsicum	340.15692	7.295	1445814.059
POS	2-Hydroxypropanoic acid; 3Z-Hexenyl ester	6.1039	DNP-Capsicum	249.03491	0.687	1369513.537
Neg	Kaempferol 3-glycosides; 3-O-[4-Hydroxy-E-cinnamoyl-(→6)-a-D-mannopyranoside]	8.0684	DNP-Solanaceae	593.13037	3.308	1349301.262
POS	Spirostane-2,3,5-triol; 3-O-[a-L-Rhamnopyranosyl-(1→2)-β-D-glucopyranoside]	7.1303	DNP-Capsicum	757.43164	4.303	1322450.061
POS	5,11-Epoxy-9-hydroxy-7-megastigmen-3-one; 9-O-β-D-Glucopyranoside	6.5198	DNP-Solanaceae	387.19769	3.432	1304578.886
Neg	Foetidinol; 3-O-[β-D-Xylopyranosyl-(1→3)-β-D-xylopyranoside]	6.988	DNP-Capsicum	707.36139	9.849	1292703.57

Tabla 64. Áreas relativas de metabolitos no volátiles de la accesión Dulce rojo

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	tr (min).	promedio de áreas
POS	Capsaicin	0.89	GNPS	306.20602	7.293	91927584
Neg	(S)-MALATE	88	all-public	133.00999	0.736	58293896
POS	Dihydrocapsaicin	84	all-public	308.22156	7.726	52183815
POS	4-Hydroxy-2-pyrrolidinecarboxylic acid; N-Me	7.1684	DNP-Capsicum	146.07973	0.707	43779285
Neg	Citric acid	87.7	all-public	191.01347	0.896	38179158
POS	Isoleucine; (2R,3S)-form	7.1426	DNP-Solanaceae	132.10056	1.02	15288708.5
POS	L-Asparagine ; LC-ESI-ITFT ; MS2 ; m/z	88	all-public	133.05943	0.719	11938156.75
POS	Glucuronic acid; 6-Aminohexyl glycoside	6.3678	DNP-Solanaceae	294.15216	1.001	11375416.25
POS	L-Phenylalanine	8.2961	Msfinder	166.08461	1.456	10384073.5
POS	Capsaicin; Nonanoyl analogue	7.6985	DNP-Capsicum	294.20645	7.178	8856828.75
POS	L-(+)-Arginine; LC-ESI-ITFT ; MS2 ; m/z	91.5	all-public	175.1172	0.669	6576857.25
Neg	6-O-?-D-Apiofuranosyl-D-glucose; D-Pyranose-form	7.9243	DNP-Capsicum	311.099	0.659	4372136.25
POS	Phosphocholine	0.91	GNPS	184.03519	0.669	4261395.75
POS	Tryptophan	88	all-public	205.09537	2.525	4017341
POS	4-Hydroxy-3-methyl-2-butenic acid; Nitrile, O-[2E-butenoyl-(?4)-?-D-glucopyranoside]	7.5954	DNP-Solanaceae	328.13608	1.456	3924556.5
POS	3,20-Epoxy-18-nor-15-beyerene-3,6-diol; (3aOH,6a)-form	7.876	DNP-Capsicum	322.23792	8.259	3756761
Neg	Foetidinol; 3-O-[?-D-Xylopyranosyl-(1?3)-?-D-xylopyranoside]	7.1091	DNP-Capsicum	707.362	9.841	3756721.25
POS	5,5"-Dicapsaicin; 6",7"-Dihydro	7.0411	DNP-Capsicum	633.38678	7.293	3544998.5
POS	Adenosine	8.0227	DNP-Solanaceae	268.10165	0.89	3235229
POS	6,13-Epoxy-4,8,9,12-eunicellanetetrol; 4-Ac	7.3038	DNP-Solanaceae	399.27304	9.834	3125727.25

Tabla 65. Continuación tabla 64...

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	tr (min).	promedio de áreas
Neg	4-Amino-4-deoxyglucose; 2,3:5,6-Di-O-isopropylidene, di-Me acetal	7.1523	DNP-Solanaceae	350.18558	7.286	3041843.5
Neg	Silychristin; 3-Deoxy	7.3555	DNP-Solanaceae	465.12439	3.958	2809797.25
POS	Uridine	0.73	GNPS	171.01497	0.676	2781920.5
Neg	3,3',4',5,7-Pentahydroxyflavan-4-acetic acid; 3'-Deoxy	7.4707	DNP-Solanaceae	313.06708	0.711	2730647.75
POS	3',4',5,7-Tetrahydroxy-3-methoxyflavone; 7-O-[α -L-Rhamnopyranosyl-(1 \rightarrow 6)-[β -D-xylopyranosyl-(1 \rightarrow 2)]- β -D-glucopyranoside]	7.5398	DNP-Solanaceae	757.21478	4.18	2632297.5
POS	2-Methoxyadenosine; 2'-Deoxy	7.0082	DNP-Solanaceae	282.11584	0.73	2524363.5
POS	Lycoclavine	6.5165	DNP-Solanaceae	330.20389	7.726	2358852.5
Neg	Acaciabiuronic acid; β -Pyranose-form	7.6599	DNP-Capsicum	355.09158	3.243	2339908.5
POS	indoline	6.9814	Msfinder	120.07957	1.456	2299688.75
POS	Shahidine	7.4462	DNP-Capsicum	280.13678	0.89	2247662.25
POS	Capsaicin; Decanoyl analogue	6.4753	DNP-Capsicum	637.41846	7.726	2122282.25
Neg	Osterivolone C	6.5734	DNP-Solanaceae	559.29321	8.564	2024269.75
POS	3,8,11,12,14-Pentahydroxypregn-5-en-20-one; 3-O-[6-Deoxy-3-O-methyl- β -D-allopyranosyl-(1 \rightarrow 4)- β -D-oleandropyranoside]	6.7486	DNP-Capsicum	685.37506	9.834	1994105.75
POS	Lincosamine; β -Pyranose-form	7.4837	DNP-Solanaceae	224.11076	0.674	1912636.25
Neg	3,16-Dihydroxyergosta-8,16-diene-15,23-dione; 3-O-[β -D-Xylopyranosyl-(1 \rightarrow 3)- β -D-glucuronopyranoside]	7.1616	DNP-Solanaceae	749.37134	10.183	1908001
POS	N-Fructosyl isoleucine	85.5	all-public	276.14172	1.001	1800226.75
Neg	3,4-Dihydroxy-2-methylenebutanoic acid; β -D-Glucopyranos-6-yl ester	7.6749	DNP-Solanaceae	293.08899	0.679	1792290.75
POS	Capsaicin; 9-Methyl-7E-decenoyl analogue	7.8606	DNP-Capsicum	320.22153	7.81	1789052.75
Neg	3-Hydroxy-1-(4-hydroxyphenyl)-5-phenyl-1-pentanone; 4'-O- β -D-Glucopyranoside	7.0582	DNP-Solanaceae	431.17755	3.441	1735550.75
POS	4-Amino-3-hydroxybenzoic acid; Me ether; Me ester	8.0294	DNP-Solanaceae	182.07959	0.896	1717243.75

Tabla 66. Áreas relativas de metabolitos no volátiles de la accesión Miscucho

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	t _R (min).	promedio de áreas
POS	Capsaicin	0.9	GNPS	306.20663	7.27	106436778
POS	Dihydrocapsaicin	84	all-public	308.22202	7.727	66770854
Neg	(S)-MALATE	88	all-public	133.0099	0.74	66059884
POS	Allysine	97	all-public	146.07962	0.705	34520295
Neg	Citric acid	87.5	all-public	191.01328	0.887	36245223
POS	2-C-Hydroxymethylribose; β-D-Pyranose-form	6.5281	DNP-Solanaceae	219.02455	0.686	24507253.5
POS	Ethyl-1,4-benzoquinone	6.6409	DNP-Solanaceae	137.05952	7.27	19857346
POS	2,5-Dimethyl-1,4-benzoquinone	7.0694	DNP-Solanaceae	137.05946	7.72	14253618
Neg	trans-Ferulic acid	7.5729	Msfinder	215.02539	0.689	15163679.25
POS	Phenylalanine; (R)-form	6.8186	DNP-Solanaceae	166.08461	1.452	12275324
POS	L- (+)-Arginine	96.2	all-public	175.11734	0.68	10860031.75
POS	Paniculine	6.2873	DNP-Solanaceae	615.43707	7.727	8692314.5
POS	L-Isoleucine	6.9856	Msfinder	132.10048	1.018	9737946.75
POS	L-Asparagine	87.2	all-public	133.05937	0.716	9752925.75
POS	Capsaicin; Nonanoyl analogue	7.8043	DNP-Capsicum	294.20645	7.173	9103528.25
POS	4-Hydroxybenzyl alcohol; 4-Me ether, α-O-(2-O-sulfo-β-D-glucopyranoside)	6.9061	DNP-Solanaceae	381.07593	0.705	7665152.75
POS	Tryptophan	88	all-public	205.09521	2.522	6123558.25
Neg	1,3,27-Trihydroxywitha-5,24-dienolide; 1-Ac, 3-O-β-D-glucopyranoside	6.9428	DNP-Solanaceae	707.36218	9.842	5565023.5
Neg	5-Hydroxy-2-(4-hydroxyphenyl) benzofuran	6.6024	DNP-Solanaceae	225.05371	1.093	4782604.75
POS	6,13-Epoxy-4,8,9,12-eunicellanetetrol; 4-Ac	7.333	DNP-Solanaceae	399.27362	9.833	4664544.25

Tabla 67. Continuación tabla 66....

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	t _R (min).	promedio de áreas
POS	3,20-Epoxy-18-nor-15-beyerene-3,6-diol; (3aOH,6a)-form	7.9358	DNP-Capsicum	322.23798	8.24	4352907.75
POS	Acetamide; N-(13-Methyltetradecyl)	7.1152	DNP-Capsicum	270.27744	10.597	3680821.5
Neg	Sigmoidin C; 8'-Deoxy, 3',4'-dihydro, 3'',4''-dihydroxy	7.4736	DNP-Solanaceae	371.10675	0.683	4017010.25
Neg	2-Deoxy-arabino-hexonic acid; D-form	7.4212	DNP-Solanaceae	179.05003	1.12	4038176.75
Neg	4-Amino-4-deoxyglucose; 2,3:5,6-Di-O-isopropylidene, di-Me acetal	7.2698	DNP-Solanaceae	350.18546	7.294	4439182.25
POS	5,5'''-Dicapsaicin; 6'',7''-Dihydro	6.9537	DNP-Capsicum	633.38751	7.27	4178197.25
Neg	2,3,5,7,8,9,15-Heptahydroxy-6(17),11-jatrophadien-14-one; 5-O-(2-Methylbutanoyl), 7-O-(2-methylpropanoyl), 2,8,9-tri-Ac	7.1048	DNP-Solanaceae	679.33124	8.971	3912213.5
POS	Lysine; 6-N-Benzoyl	7.7719	DNP-Solanaceae	251.1393	2.035	3947048.75
Neg	3-Hydroxy-1-(4-hydroxyphenyl)-5-phenyl-1-pentanone; 4'-O-β-D-Glucopyranoside	6.957	DNP-Solanaceae	431.17722	3.436	3497206
Neg	3-Phenyl-2-propen-1-ol; O-[α-L-Rhamnopyranosyl-(1 → 6)-β-D-glucopyranoside]	7.5967	DNP-Capsicum	441.18253	4.077	3389155
POS	Capsaicin; Decanoyl analogue	6.3588	DNP-Capsicum	330.2041	7.717	2854430.5
POS	6-Dodecyl-3-hydroxy-2-methylpyridine; 7'-Oxo, 12'-hydroxy	6.2325	DNP-Solanaceae	637.4187	7.72	2875727.5
Neg	4(20),11-Taxadiene-2,5,10,14-tetrol; 14-O-(3S-Hydroxy-2R-methylbutanoyl), 2,5-di-Ac, 10-O-β-D-glucopyranoside	7.1381	DNP-Solanaceae	681.34668	9.648	3176559.5
Neg	3,11,15,24,25-Pentahydroxycycloart-7-ene-16,23-dione; 3-O-β-D-Xylopyranoside	7.1875	DNP-Capsicum	695.36224	10.008	3514413.75
POS	L-Phenylalanine	0.93	GNPS	120.0795	1.446	2821255.75
Neg	6-O-β-D-Apiofuranosyl-D-glucose; D-Pyranose-form	7.9047	DNP-Capsicum	311.09888	0.663	2621379
Neg	8-Formyl-2,5,7-trihydroxy-6-methylflavanone	7.4392	DNP-Solanaceae	313.06689	0.721	2708186.75
POS	2-Oxobutanoic acid; Diethylamide	8.1654	DNP-Solanaceae	158.11595	0.723	2311246.5
POS	6-Nonyl-2-cyclohexen-1-ol; O-β-D-Glucopyranoside	7.433	DNP-Solanaceae	387.27319	9.988	2507474.5
POS	aerugine	7.3723	Mfinder	210.05891	0.896	2632336.75

Tabla 68. Áreas relativas de metabolitos no volátiles de la accesión Tomatito rojo

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Dbs	Mz	tr (min).	promedio de áreas
POS	Capsaicin	0.9	GNPS	306.20679	7.278	155047784
Neg	Citric acid	88	all-public	191.01344	0.74	89728878
POS	5-Hydroxy-2-piperidinecarboxylic acid; (2S,5S)-form	6.541	DNP-Solanaceae	146.07962	0.707	34851838
POS	Dihydrocapsaicin	83.9	all-public	308.22189	7.732	23102981.5
Neg	(S)-MALATE	88	all-public	133.00999	0.734	17779463.5
Neg	2,3-Dihydro-3,5-dihydroxy-2-methyl-4H-1-benzopyran-4-one; (2R*,3R*)-form	6.5301	DNP-Solanaceae	215.02554	0.682	16636669.25
POS	1,2,3,4-Tetrahydro-7,8-dihydroxyisoquinoline	7.4505	DNP-Solanaceae	166.08473	1.447	13067644
Neg	NCGC00385770-01!2-(3,4-dihydroxyphenyl)-5-hydroxy-7-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[(3,4,5-trihydroxyoxan-2-yl)oxymethyl]oxan-2-yl]oxochromen-4-one	0.75	GNPS	579.11572	3.908	7126922.75
POS	Isoleucine; (2R,3S)-form	7.7199	DNP-Solanaceae	132.10051	1.027	7005907.25
POS	Capsaicin; Nonanoyl analogue	7.9143	DNP-Capsicum	294.20648	7.183	6332830.5
POS	Glucuronic acid; 6-Aminoethyl glycoside	7.353	DNP-Solanaceae	294.15219	1.006	6009507.75
POS	4-Hydroxy-3-methyl-2-butenic acid; Nitrile, O-[2E-butenoyl-(?)-β-D-glucopyranoside]	7.586	DNP-Solanaceae	328.13623	1.437	5332217.5
POS	Tryptophan	88	all-public	205.09641	2.515	4934648.25
Neg	2-Deoxy-arabino-hexonic acid	7.8381	DNP-Solanaceae	225.05392	1.019	4887807.75
POS	2,5-Dimethyl-1,4-benzoquinone	7.4743	DNP-Solanaceae	137.05942	7.732	4824339.75
POS	Ethyl-1,4-benzoquinone	7.717	DNP-Solanaceae	137.05959	7.278	4495807
Neg	Isobar: glucose, fructose, mannose, galactose	0.92	GNPS	179.05019	1.019	4011198
Neg	quercitrin	0.91	GNPS	447.07809	4.227	3885888.25
POS	L-Asparagine	85.6	all-public	133.0594	0.719	3647761
Neg	Dehydroascorbic acid	8.7769	Msfinder	173.00346	1.225	3482338.25

Tabla 69. Continuación tabla 68...

Modo de ionización	Nombre del metabolito	Dot product/score/cosine	Db	Mz	t _R (min).	promedio de áreas
Neg	5-Hydroxy-2-(4-hydroxyphenyl) benzofuran	6.7874	DNP-Solanaceae	225.05397	1.331	3352386
POS	Phosphocholine	0.92	GNPS	184.03525	0.663	3321095.5
Neg	6-O-a-D-Glucopyranuronosyl-D-glucose; a-Pyranose-form	7.5813	DNP-Capsicum	355.09143	3.242	3234121.75
Neg	Coumaroyl Hexoside (isomer of 690, 692); PlaSMA ID-691	85.1	all-public	325.08206	3.034	3110371
POS	L-Phenylalanine	0.91	GNPS	120.07994	1.447	2937992
POS	Quercetin-3-O-rhamnoside	90.9	all-public	449.10376	4.213	2812461.25
POS	3-(4-Hydroxyphenyl)-2-propenoic acid; 4'-O-Sulfate	6.0919	DNP-Capsicum	245.00375	0.75	2688290.75
Neg	Luteolin-6-C-glucoside	0.87	GNPS	447.07791	3.558	2083671.5
POS	3,3',5,7-Tetrahydroxy-4'-methoxyflavone; 3-O-Arabinoside	7.8281	DNP-Solanaceae	449.1037	3.63	2024932
POS	L- (+)-Arginine	92.3	all-public	175.11719	0.676	1807458.5
Neg	2,3,14,20,22,28-Hexahydroxy-6-oxostigmast-7-en-26-oic acid; 26?28-Lactone, 22-Ac	7.5088	DNP-Solanaceae	561.30847	9.128	1732532.75
POS	3,20-Epoxy-18-nor-15-beyerene-3,6-diol; (3aOH,6a)-form	7.9454	DNP-Capsicum	322.23792	8.257	1675038.75
POS	Abrine	0.89	GNPS	188.06888	2.515	1595710.5
POS	3-O-a-D-Mannopyranosyl-D-glucose;β-D-Pyranose-form	7.8483	DNP-Capsicum	325.11038	0.756	1588013.75
Neg	Kaempferol 7-glycosides; 7-O-[2-Methoxybenzoyl-(→3)-a-L-rhamnopyranoside]	7.3794	DNP-Capsicum	565.13678	4.227	1506014.25
POS	Spirostane-2,3,5-triol; 3-O-[a-L-Rhamnopyranosyl-(1→4)-β-D-glucopyranoside]	7.5328	DNP-Solanaceae	757.43164	4.324	1431104
POS	Capsaicin; 9-Methyl-7E-decenoyl analogue	7.622	DNP-Capsicum	320.22159	7.806	1374953
Neg	Affinoside A; 16,17-Didehydro	6.6195	DNP-Solanaceae	571.22321	4.626	1320495
POS	1H-Pyrrole-2-carboxylic acid; ?-D-Glucopyranosyl ester	7.2952	DNP-Solanaceae	274.08972	0.895	1319864.25

13.2.ANEXO 2 (Perfil de Metabolitos Volátiles)

Tabla 70. Áreas relativas de metabolitos volátiles de la accesión Ayuello, (ají fresco)

Nombre del metabolito	Dot product	dfs	tr (min).	índice de Kovats	promedio de áreas
beta-Caryophyllene	99.5	flavor	44.59	1428.37	154766650
N,N,O-Triacetylhydroxylamine	100	replib	37.626	1316.11	74146590
Ethanethioic acid, S-(2-methylbutyl) ester	90	mainlib	31.319	1214.79	71656695
1-Penten-3-one, 2-methyl-	88.1	replib	12.284	849.55	64876235
4-Methylpentyl 4-methylpentanoate	93.8	mainlib	39.86	1351.9	64546360
Tridecane, 2-methyl-	99.2	replib	40.683	1365.09	62788145
cis-Calamenene	99	mainlib	50.458	1526.48	56493905
2(3H)-Furanone, 5-ethylidihydro-	79.4	mainlib	30.771	1205.97	56477745
(S)-(-)-3-Methyl-1-pentanol	97.2	mainlib	12.701	860.13	53620980
Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	98.7	replib	50.266	1523.11	48447515
n-Butyl ether	92	replib	12.612	857.86	45310460
Pentanoic acid, 4-hexen-1-yl ester	94.3	mainlib	32.636	1235.96	44196245
1-Hexene, 3,5-dimethyl-	90.7	mainlib	12.252	848.73	36338490
E-2-Hexadecacen-1-ol	94.7	mainlib	45.697	1446.54	34508070
2-Methylpiperidine	97.9	replib	37.306	1310.99	33684609.45
1H-Indene, 1-methylene-	93.8	mainlib	29.876	1191.21	31510440
Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	99.2	replib	42.496	1394.12	31359165
Oxa-cycloundec-8-ene-2,7-dione	85.5	mainlib	39.553	1346.98	28930675
cis-3-Hexenyl isovalerate	96.5	flavor	32.559	1234.73	28503775
Isoamyl isovalerate	99.4	flavor	24.745	1105.03	26653690

Tabla 71. Continuación Tabla 70....

Nombre del metabolito	Dot product	dbs	t _R (min)	índice de Kovats	promedio de áreas
Isoamyl isovalerate	99.4	flavor	24.745	1105.03	26653690
Oxacyclotetradecan-2-one, 14-methyl-	97.9	mainlib	56.742	1662.7	25982000
Pentadecane	89.8	flavor	48.972	1500.38	24455305
Hexadecane, 2,6,11,15-tetramethyl-	86	replib	46.814	1464.9	21543975
alfa-Copaene	98.6	mainlib	41.655	1380.66	18640600
Naphthalene, 1,2,3,4,4a,5,6,8a-octahydro-7-methyl-4-methylene-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)-	99.6	replib	49.989	1518.24	17523080
Pentanoic acid, 4-methyl-, pentyl ester	94.6	mainlib	33.67	1252.58	14009855
Heptanoic acid, 3-methylbutyl ester	85.7	mainlib	37.88	1320.19	13968935
Butanoic acid, 3-methyl-, hexyl ester	99.6	replib	33.001	1241.83	13889850
Hexanoic acid, 3-hexenyl ester, (Z)-	98.7	replib	41.59	1379.62	12870116.5
Bicyclo [4.4.0] dec-1-ene, 2-isopropyl-5-methyl-9-methylene-	97.3	replib	47.006	1468.04	12320940
2-Hexen-1-ol (trans)	99.7	flavor	12.586	857.21	11763995
Isoamyl hexanoate	99.9	flavor	33.479	1249.51	11463648.5
1H-Cycloprop[e]azulene, 1a,2,3,4,4 a,5,6,7b-octahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.alpha.,4a.beta.,7b.alpha.)]-	96.4	mainlib	49.09	1502.43	11043615
Butanoic acid,4-hexen-1-yl ester	99.2	mainlib	12.232	848.23	10909243.5
4-Methylpentyl 2-methylbutanoate	97.3	mainlib	30.388	1199.78	10849876.5
Bicyclosesquiphellandrene	98.4	mainlib	42.41	1392.75	10787454.5
gamma-muurolene	92.4	All-public	47.72	1479.77	10134678.5
Heptanoic acid, 3-hexenyl ester, (Z)-	86.4	mainlib	45.882	1449.59	10071923.5
.alpha.-Cubebene	96.7	replib	39.827	1351.38	9870613

Tabla 72. Áreas relativas de metabolitos volátiles de la accesión Charapita, (ají fresco)

Nombre del metabolito	Dot product	dbs	t _R (min)	índice de Kovats	promedio de áreas
Butanoic acid, 3-methyl-, heptyl ester	99.8	mainlib	39.153	1340.58	134212700
Butanoic acid, 3-methyl-, hexyl ester	99.6	replib	33.001	1241.83	109135000
2(3H)-Furanone, 5-ethylidihydro-	79.4	mainlib	30.771	1205.97	100647430.9
4-Methylpentyl 2-methylbutanoate	97.3	mainlib	30.388	1199.78	98753505
cis-Calamenene	99	mainlib	50.458	1526.48	84688205
citronellyl valerate	93.5	All-public	53.73	1584.02	76186740
Butyl heptanoate	80.9	flavor	31.672	1220.46	72018980
.alpha.-ylangene	90.3	mainlib	41.813	1383.19	71353635
Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	98.7	replib	50.266	1523.11	70246965
Butanoic acid, 2-methyl-, heptyl ester	99.5	replib	38.724	1333.7	70205230
.alpha.-Cubebene	96.7	replib	39.827	1351.38	69602205
Hexadecane, 2,6,11,15-tetramethyl-	86	replib	46.814	1464.9	68991875
Hexanoic acid, 3-hexenyl ester, (Z)-	98.7	replib	41.59	1379.62	66885955
Tridecane, 2-methyl-	99.2	replib	40.683	1365.09	65766885
6-Methylheptyl 3-methylbutanoate	87.2	mainlib	43.051	1403.11	65459660
6-Methylhept-4-en-1-yl isobutyrate	98.3	mainlib	36.03	1290.52	59145040
Pentanoic acid, pentyl ester	90.4	mainlib	29.848	1190.75	49311886.5
1H-Indene, 1-methylene-	93.8	mainlib	29.876	1191.21	48793540
Pentanoic acid, 4-hexen-1-yl ester	94.3	mainlib	32.636	1235.96	46578485
5-Isopropenyl-2-methyl-7-oxabicyclo[4.1.0]heptan-2-ol	84.1	mainlib	33.155	1244.31	46470735

Tabla 73. Continuación Tabla 72...

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
E-2-Hexadecacen-1-ol	94.7	mainlib	45.697	1446.54	43514305
Methylsalicylate	92.9	flavor	30.193	1196.53	43114740
trans-2-Hexenal	97.7	flavor	11.832	838.07	41546095
cis-3-Hexenyl isovalerate	96.5	flavor	32.559	1234.73	41374745
Octyl isovalerate	99.6	flavor	45.327	1440.47	41326265
Naphthalene, 1,7-dimethyl-	87	replib	44.928	1433.93	41326265
Citronellyl propionate	82.9	flavor	53.477	1579.57	41127885
4-Methylpentyl 4-methylpentanoate	93.8	mainlib	39.86	1351.9	37768790
5-Methylhexyl 3-methylbutanoate	89.1	mainlib	37.21	1309.45	37610960
N, N, O-Triacetylhydroxylamine	100	replib	37.626	1316.11	33260270
Oxalic acid, diallyl ester	92.7	mainlib	46.449	1458.9	31538192.8
Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	99.2	replib	42.496	1394.12	28205330
(S)-(-)-3-Methyl-1-pentanol	97.2	mainlib	12.701	860.13	28177745
Propanoic acid, 2-methyl-, 4-methylpentyl ester	98.4	mainlib	25.035	1109.9	27797035
gamma-muurolene	92.4	All-public	47.72	1479.77	27243085
6-Methyl-4-heptenyl pentanoate	86.9	mainlib	44.937	1434.08	26584420
Naphthalene, 1,2,3,4,4a,5,6,8a-octahydro-7-methyl-4-methylene-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)-	99.6	replib	49.989	1518.24	26362415
Bicyclosesquiphellandrene	98.4	mainlib	42.41	1392.75	25383710
Propanoic acid, 2-methyl-, heptyl ester	93.8	mainlib	33.221	1245.37	25137505
1-Isopropyl-4,7-dimethyl-1,2,3,4,5,6-hexahydronaphthalene	91.5	mainlib	47.549	1476.96	25089915

Tabla 74. Áreas relativas de metabolitos volátiles de la accesión Dulce rojo, (ají fresco)

Nombre del metabolito	Dot product	dbs	tr (min)	índice de Kovats	promedio de áreas
N,N,O-Triacetylhydroxylamine	100	replib	37.626	1316.11	103590965
Methylsalicylate	92.9	flavor	30.193	1196.53	65025680
2(3H)-Furanone, 5-ethylidihydro-	79.4	mainlib	30.771	1205.97	57380320
4-Methylpentyl 4-methylpentanoate	93.8	mainlib	39.86	1351.9	46862480
trans-2-Hexenal	97.7	flavor	11.832	838.07	28592660
4-Methylpentyl 2-methylbutanoate	97.3	mainlib	30.388	1199.78	24713300
cis-Calamenene	99	mainlib	50.458	1526.48	23638080
Oxa-cycloundec-8-ene-2,7-dione	85.5	mainlib	39.553	1346.98	19711287.5
cis-3-Hexenyl isovalerate	96.5	flavor	32.559	1234.73	18871665
Cyclopentanecarboxylic acid, hexyl ester	87.5	mainlib	40.529	1362.62	18523395
(3E,7E)-4,8,12-Trimethyltrideca-1,3,7,11-tetraene	95.5	mainlib	53.217	1574.99	15621955
1-Penten-3-one, 2-methyl-	88.1	replib	12.284	849.55	15189420.5
Butanoic acid, 3-methyl-, hexyl ester	99.6	replib	33.001	1241.83	14705975
.alpha.-Cubebene	96.7	replib	39.827	1351.38	14001520
1H-Indene, 1-methylene-	93.8	mainlib	29.876	1191.21	12942585
Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	98.7	replib	50.266	1523.11	11228180
Germacrene D	97.3	replib	48.178	1487.29	10801460
cis-3-Hexenyl-2-methyl-butanoate	96.7	flavor	32.273	1230.12	8967949.5
4-Aminobutanoic acid	92.9	mainlib	33.643	1252.15	8073781
alfa-Copaene	98.6	mainlib	41.655	1380.66	7711232

Tabla 75. Continuación Tabla 74...

Nombre del metabolito	Dot product	dbs	tr (min)	índice Kovats	de	promedio áreas	de
1-Hexene, 3,5-dimethyl-	90.7	mainlib	12.252	848.73		7703549	
Hexadecane, 2,6,11,15-tetramethyl-	86	replib	46.814	1464.9		7237749	
5-Methylhexyl 3-methylbutanoate	89.1	mainlib	37.21	1309.45		7197013	
n-Butyl ether	92	replib	12.612	857.86		7127233.5	
Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	99.2	replib	42.496	1394.12		6766477	
12-Methyl-oxa-cyclododecan-2-one	88.7	mainlib	47.147	1470.37		6087158.5	
E-2-Hexadecacen-1-ol	94.7	mainlib	45.697	1446.54		5996887	
Heptanoic acid, 3-methylbutyl ester	85.7	mainlib	37.88	1320.19		5705712.95	
Tridecane, 2-methyl-	99.2	replib	40.683	1365.09		5288325.5	
Bicyclosesquiphellandrene	98.4	mainlib	42.41	1392.75		5235154.5	
Pentanoic acid, 4-methyl-, pentyl ester	94.6	mainlib	33.67	1252.58		4965007.5	
Ethanethioic acid, S-(2-methylbutyl) ester	90	mainlib	31.319	1214.79		4736438.5	
Pentadecane	89.8	flavor	48.972	1500.38		4406076.5	
(S)-(-)-3-Methyl-1-pentanol	97.2	mainlib	12.701	860.13		4313467.5	
Butanoic acid,4-hexen-1-yl ester	99.2	mainlib	12.232	848.23		4302717.5	
2-Hexen-1-ol (trans)	99.7	flavor	12.586	857.21		4049512	
Pyrazine, 2-methoxy-3-(2-methylpropyl)-	99.8	mainlib	28.949	1175.65		3511563.5	
Bicyclo[4.4.0]dec-1-ene, 2-isopropyl-5-methyl-9-methylene-	97.3	replib	47.006	1468.04		3435147.5	
1-Pentanol, 4-methyl-	96.7	mainlib	11.297	824.5		3334291.5	
gamma-muurolene	92.4	All-public	47.72	1479.77		3132892	

Tabla 76. Áreas relativas de metabolitos volátiles de la accesión Miscucho, (ají fresco)

Nombre del metabolito	Dot product	dbs	tr (min)	índice de Kovats	promedio de áreas
N,N,O-Triacetylhydroxylamine	100	replib	37.626	1316.11	149088250
4-Methylpentyl 4-methylpentanoate	93.8	mainlib	39.86	1351.9	94642310
Cyclopentanecarboxylic acid, hexyl ester	87.5	mainlib	40.529	1362.62	52836015
Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, [1S-(1.alpha.,4a.beta.,8a.alpha.)]-	98.7	replib	50.266	1523.11	51055175
Hexadecane, 2,6,11,15-tetramethyl-	86	replib	46.814	1464.9	48903160
cis-Calamenene	99	mainlib	50.458	1526.48	45564935
E-2-Hexadecacen-1-ol	94.7	mainlib	45.697	1446.54	43435180
1H-Indene, 1-methylene-	93.8	mainlib	29.876	1191.21	41243935
.alpha.-Cubebene	96.7	replib	39.827	1351.38	36171305
(S)-(-)-3-Methyl-1-pentanol	97.2	mainlib	12.701	860.13	35649045
alfa-Copaene	98.6	mainlib	41.655	1380.66	34361290
2(3H)-Furanone, 5-ethylidihydro-	79.4	mainlib	30.771	1205.97	29103185
Oxa-cycloundec-8-ene-2,7-dione	85.5	mainlib	39.553	1346.98	27693850
trans-2-Hexenal	97.7	flavor	11.832	838.07	24830065
Germacrene D	97.3	replib	48.178	1487.29	23904260
Tridecane, 2-methyl-	99.2	replib	40.683	1365.09	23235410
Pentadecane	89.8	flavor	48.972	1500.38	17429505
gamma-muurolene	92.4	All-public	47.72	1479.77	16549030
Methylsalicylate	92.9	flavor	30.193	1196.53	16438395
Naphthalene, 1,2,3,4,4a,5,6,8a-octahydro-7-methyl-4-methylene-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)-	99.6	replib	49.989	1518.24	15996710

Tabla 77. Continuación Tabla 76...

Nombre del metabolito	Dot product	dbs	tr (min)	índice de Kovats	promedio de áreas
1-Isopropyl-4,7-dimethyl-1,2,3,4,5,6-hexahydronaphthalene	91.5	mainlib	47.549	1476.96	15602505
Cyclopropane, 1-methyl-1-(1-methylethyl)-2-nonyl-	92.4	mainlib	51.275	1540.84	15304135
Cubenene	93	mainlib	51.103	1537.83	14854250
1H-Cycloprop[e]azulene, 1a,2,3,4,4a,5,6,7b-octahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.alpha.,4a.beta.,7b.alpha.)]-	96.4	mainlib	49.09	1502.43	13910195
4-Methylpentyl 8-methylnon-6-enoate	98.5	mainlib	57.647	1689.71	13888844
Butanoic acid,4-hexen-1-yl ester	99.2	mainlib	12.232	848.23	12770385
(3E,7E)-4,8,12-Trimethyltrideca-1,3,7,11-tetraene	95.5	mainlib	53.217	1574.99	12766013
Butanoic acid, 3-methyl-, hexyl ester	99.6	replib	33.001	1241.83	12491240
Pyrazine, 2-methoxy-3-(2-methylpropyl)-	99.8	mainlib	28.949	1175.65	12316205
4-Isopropyl-6-methyl-1-methylene-1,2,3,4-tetrahydronaphthalene	97.8	mainlib	51.626	1547.02	12050335
Bicyclosesquiphellandrene	98.4	mainlib	42.41	1392.75	12029830
4-Methylpentyl 2-methylbutanoate	97.3	mainlib	30.388	1199.78	11792024
Bicyclo[4.4.0]dec-1-ene, 2-isopropyl-5-methyl-9-methylene-	97.3	replib	47.006	1468.04	11762620
1-Pentanol, 4-methyl-	96.7	mainlib	11.297	824.5	11708903
1-Penten-3-one, 2-methyl-citronellyl valerate	88.1	replib	12.284	849.55	11022991.5
	93.5	All-public	53.73	1584.02	10488993
Cyclohexane, 1,1,3-trimethyl-	91.2	mainlib	18.608	990.91	10453846
Butanoic acid, 3-methyl-, heptyl ester	99.8	mainlib	39.153	1340.58	9581490.5
Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, [1S-(1.alpha.,2.beta.,4.beta.)]-	99.2	replib	42.496	1394.12	9360135.5
4-Aminobutanoic acid	92.9	mainlib	33.643	1252.15	9113263

Tabla 78. Áreas relativas de metabolitos volátiles de la accesión Tomatito rojo, (ají fresco)

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
(S)-(-)-3-Methyl-1-pentanol	97.2	mainlib	12.701	860.13	77208620
n-Butyl ether	92	replib	12.612	857.86	76058840
1-Penten-3-one, 2-methyl-	88.1	replib	12.284	849.55	67697310
Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-, (1.alpha.,2.beta.,4.beta.)]-	99.2	replib	42.496	1394.12	63628883.45
cis-Calamenene	99	mainlib	50.458	1526.48	57326193.33
4-Methylpentyl 4-methylpentanoate	93.8	mainlib	39.86	1351.9	50212840
1-Hexene, 3,5-dimethyl-	90.7	mainlib	12.252	848.73	45470886.67
Bicyclosesquiphellandrene	98.4	mainlib	42.41	1392.75	41035919.33
Naphthalene, 1,2,4a,5,8,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)]-	98.7	replib	50.266	1523.11	40657080
Oxacyclotetradecan-2-one, 14-methyl-	97.9	mainlib	56.742	1662.7	38711553.33
alfa-Copaene	98.6	mainlib	41.655	1380.66	31915943.33
Butanoic acid, 3-methyl-, hexyl ester	99.6	replib	33.001	1241.83	27185350
Oxa-cycloundec-8-ene-2,7-dione	85.5	mainlib	39.553	1346.98	22933793.67
1H-Indene, 1-methylene-	93.8	mainlib	29.876	1191.21	21999180
beta-Humulene	93.3	mainlib	48.63	1494.72	20784156
1H-Cycloprop[e]azulene, 1a,2,3,4, 4a,5,6,7b-octahydro-1,1,4,7-tetramethyl-, [1aR-(1a.alpha.,4.alpha.,4a.beta.,7b.alpha.)]-	96.4	mainlib	49.09	1502.43	20222690.33
Ethanethioic acid, S-(2-methylbutyl) ester	90	mainlib	31.319	1214.79	19788816.67
N,N,O-Triacetylhydroxylamine	100	replib	37.626	1316.11	17930247.17
Butanoic acid,4-hexen-1-yl ester	99.2	mainlib	12.232	848.23	17231966.67
beta-Caryophyllene	99.5	flavor	44.59	1428.37	15384180

Tabla 79. Continuación Tabla 78...

Nombre del metabolito	Dot product	db	tr (min)	índice de Kovats	promedio de áreas
2-Hexen-1-ol (trans)	99.7	flavor	12.586	857.21	14743559.67
3-Hexenal	98.9	mainlib	9.901	789.86	14538042.67
2(3H)-Furanone, 5-ethylidihydro-	79.4	mainlib	30.771	1205.97	13764463.67
trans,trans-2,4-Hexadienal	99.7	flavor	14.694	908.84	12424451.67
cis-3-Hexenyl isovalerate	96.5	flavor	32.559	1234.73	10984728.33
Guaiyl acetate	85.9	flavor	47.651	1478.65	10288475.33
Methylsalicylate	92.9	flavor	30.193	1196.53	10036941.67
.alpha.-Cubebene	96.7	replib	39.827	1351.38	10015747
12-Methyl-oxa-cyclododecan-2-one	88.7	mainlib	47.147	1470.37	9746663.667
1-Pentanol, 4-methyl-	96.7	mainlib	11.297	824.5	8594187.667
gamma-murolene	92.4	All-public	47.72	1479.77	8318116.333
Naphthalene, 1,2,3,4,4a,5,6,8a-octahydro-7-methyl-4-methylene-1-(1-methylethyl)-, (1.alpha.,4a.beta.,8a.alpha.)-	99.6	replib	49.989	1518.24	7506747.667
Spiro[5.5]undec-2-ene, 3,7,7-trimethyl-11-methylene-, (-)-.alpha.-ylangene	95.9	mainlib	49.113	1502.87	7472944.667
Hexyl hexanoate	90.3	mainlib	41.813	1383.19	7335367
4-Isopropyl-6-methyl-1-methylene-1,2,3,4-tetrahydronaphthalene	94.1	flavor	41.996	1386.12	7025999.267
Bicyclo[4.4.0]dec-1-ene, 2-isopropyl-5-methyl-9-methylene-	97.8	mainlib	51.626	1547.02	6976674.667
1-Isopropyl-4,7-dimethyl-1,2,3,4,5,6-hexahydronaphthalene	97.3	replib	47.006	1468.04	6767207.667
Isoamyl isovalerate	91.5	mainlib	47.549	1476.96	5648555.667
Hexane	99.4	flavor	24.745	1105.03	5575328.333
	92.6	flavor	12.372	851.78	5544297.13

Tabla 80. Áreas relativas de metabolitos volátiles de la accesión Ayuclo, (ají liofilizado)

Nombre del metabolito	Dot product	dbs	tr (min)	índice de Kovats	promedio de áreas
Benzaldehyde	96.8	replib	17.052	958.28	119083200
Terpinolene	93.4	flavor	23.571	1083.81	60805375
o-Cymene	92.9	mainlib	20.245	1022.26	60345727.5
Butyl butyrate	97.7	flavor	18.552	989.72	49604282.75
2(3H)-Furanone, 5-ethylidihydro-	85.2	mainlib	30.414	1200.24	45264890
Sebacic acid, di(2-chlorophenethyl) ester	100	mainlib	21.971	1054.2	35028712.5
Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-delta-3-carene	99.5	mainlib	15.228	920.04	28511725
beta-Pinene	98.1	flavor	15.649	928.87	26849120
beta-Pinene	99.7	flavor	17.877	975.58	26076242.5
Cyclohexane, 1-methylene-4-(1-methylethenyl)-	98.4	mainlib	19.175	1002.47	25516670
Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	91.6	replib	18.3	984.45	24489540
o-Isopropenyltoluene	97.6	mainlib	23.794	1087.95	23445277.5
Heptanoic acid, methyl ester	98.2	replib	19.987	1017.49	22311242.5
4-Decen-6-yne, (Z)-	84.5	mainlib	17.58	969.35	21429950
gamma-Terpinene	99.6	null	21.986	1054.48	21050477.5
alpha-terpinene	99.6	flavor	19.8	1014.03	18495222.5
4-Methylpentyl 2-methylbutanoate	99.5	mainlib	30.02	1193.64	12780492.5
4-Hexen-3-one	83.2	replib	20.584	1028.53	11099438.5
Limonene	95.6	replib	20.544	1027.8	10458245.75
Caryophyllene	99.9	mainlib	44.053	1419.55	10076141

Tabla 81. Continuación Tabla 80...

Nombre del metabolito	Dot product	dfs	Average tr (min)	índice de Kovats	promedio de áreas
Isoamyl hexanoate	91.4	flavor	30.978	1209.29	9370822.25
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	8871128.25
Humulene	98.9	replib	46.222	1455.17	8650623.75
Eucalyptol	99.6	mainlib	20.767	1031.91	7310108.25
Hexenal	99.8	flavor	9.791	787.22	7153579.25
Isoamyl isovalerate	97.9	flavor	24.481	1100.6	7143346.75
Phenethyl alcohol	94.5	flavor	8.783	763.4	6389730.75
Propanoic acid, 2-methyl-, 4-methylpentyl ester	99.8	mainlib	24.755	1105.21	5840832
4-Methylpentyl 4-methylpentanoate	97.7	mainlib	37.122	1307.91	5698318.5
Styrene	99.9	mainlib	13.65	884.2	5511411
Pentanoic acid, cyclohexyl ester	89.2	replib	32.216	1229.17	5055561
1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	97.9	replib	21.203	1039.99	4905629
Ethyl butyrate	99.8	flavor	9.738	785.96	4581835.25
6-Methylhept-4-en-1-yl 3-methylbutanoate	99.5	mainlib	41.571	1379.28	4065154.75
Azulene	94.8	mainlib	29.589	1186.39	3922440.75
Methyl caproate; methyl hexanoate	99.4	flavor	14.944	914.08	3842842.5
Butanoic acid, 3-methyl-, heptyl ester	87.4	mainlib	38.702	1333.26	3689419.5
(+)-2-Bornanone	99.9	replib	27.273	1147.5	3538831.25
5,5-Dimethyl-1-vinylbicyclo[2.1.1]hexane	85	mainlib	17.976	977.66	3045641.25
7-Propylidene-bicyclo[4.1.0]heptane	92.7	mainlib	18.089	980.02	2891241.5

Tabla 82. Áreas relativas de metabolitos volátiles de la accesión Charapita, (ají liofilizado)

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Benzaldehyde	96.8	replib	17.052	958.28	1.33E+08
Butyl butyrate	97.7	flavor	18.552	989.72	8.30E+07
o-Cymene	92.9	mainlib	20.245	1022.26	64366447.5
Terpinolene	93.4	flavor	23.571	1083.81	4.51E+07
4-Methylpentyl 2-methylbutanoate	99.5	mainlib	30.02	1193.64	4.35E+07
Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	91.6	replib	18.3	984.45	3.86E+07
2(3H)-Furanone, 5-ethylidihydro-	85.2	mainlib	30.414	1200.24	3.67E+07
Sebacic acid, di(2-chlorophenethyl) ester	100	mainlib	21.971	1054.2	3.52E+07
Cyclohexane, 1-methylene-4-(1-methylethenyl)-	98.4	mainlib	19.175	1002.47	3.00E+07
Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	99.5	mainlib	15.228	920.04	29812730
6-Methylhept-4-en-1-yl 3-methylbutanoate	99.5	mainlib	41.571	1379.28	2.84E+07
beta-Pinene	99.7	flavor	17.877	975.58	2.80E+07
delta-3-carene	98.1	flavor	15.649	928.87	2.71E+07
4-Hexen-3-one	83.2	replib	20.584	1028.53	2.70E+07
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	2.57E+07
Heptanoic acid, methyl ester	98.2	replib	19.987	1017.49	2.53E+07
o-Isopropenyltoluene	97.6	mainlib	23.794	1087.95	2.40E+07
alpha-terpinene	99.6	flavor	19.8	1014.03	2.33E+07
4-Decen-6-yne, (Z)-	84.5	mainlib	17.58	969.35	2.20E+07
gamma-Terpinene	99.6	null	21.986	1054.48	2.12E+07

Tabla 83. Continuación Tabla 82...

Nombre del metabolito	Dot product	dbs	tr (min)	índice de Kovats	promedio de áreas
Butanoic acid, 3-methyl-, heptyl ester	87.4	mainlib	38.702	1333.26	1.80E+07
Limonene	95.6	replib	20.544	1027.8	1.42E+07
6-Methylhept-4-en-1-yl 2-methylbutanoate	99.2	mainlib	41.179	1372.99	1.30E+07
Propanoic acid, 2-methyl-, 4-methylpentyl ester	99.8	mainlib	24.755	1105.21	1.26E+07
Pentanoic acid, cyclohexyl ester	89.2	replib	32.216	1229.17	1.20E+07
Phenethyl alcohol	94.5	flavor	8.783	763.4	1.11E+07
Isovaleric acid	93.5	flavor	11.019	817.43	10466343.83
Isoamyl isovalerate	97.9	flavor	24.481	1100.6	9375566
6-Methylheptyl 3-methylbutanoate	99.5	mainlib	42.557	1395.1	8738692
Eucalyptol	99.6	mainlib	20.767	1031.91	8.68E+06
4-Methylpentyl 4-methylpentanoate	97.7	mainlib	37.122	1307.91	7.60E+06
1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	97.9	replib	21.203	1039.99	7526361.25
5,5-Dimethyl-1-vinylbicyclo[2.1.1]hexane	85	mainlib	17.976	977.66	7390916.75
6-Methylhept-4-en-1-yl isobutyrate	99.7	mainlib	35.648	1284.25	7.17E+06
Styrene	99.9	mainlib	13.65	884.2	7062202.5
Formaldehyde oxime trimer	87.6	mainlib	18.58	990.3	6822757.5
2,6-Dimethyl-2-trans-6-octadiene	79.6	mainlib	18.728	993.42	6713728
Butanoic acid, 2-methyl-, heptyl ester	99.8	replib	38.315	1327.05	5936749.75
Methyl caproate; methyl hexanoate	99.4	flavor	14.944	914.08	5705812.5
Isoamyl hexanoate	91.4	flavor	30.978	1209.29	5485407.5

Tabla 84. Áreas relativas de metabolitos volátiles de la accesión Dulce rojo, (ají liofilizado)

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Benzaldehyde	96.8	replib	17.052	958.28	114831325
Butyl butyrate	97.7	flavor	18.552	989.72	94326735
2(3H)-Furanone, 5-ethylidihydro-	85.2	mainlib	30.414	1200.24	68645112.5
o-Cymene	92.9	mainlib	20.245	1022.26	65375683
Terpinolene	93.4	flavor	23.571	1083.81	49794140
Sebacic acid, di(2-chlorophenethyl) ester	100	mainlib	21.971	1054.2	45407357.5
beta-Pinene	99.7	flavor	17.877	975.58	41670205
delta-3-carene	98.1	flavor	15.649	928.87	41074300
4-Decen-6-yne, (Z)-	84.5	mainlib	17.58	969.35	38281682.5
gamma-Terpinene	99.6	null	21.986	1054.48	34157800
alpha-terpinene	99.6	flavor	19.8	1014.03	33570912.5
4-Methylpentyl 2-methylbutanoate	99.5	mainlib	30.02	1193.64	32155897.5
6-Methylhept-4-en-1-yl 3-methylbutanoate	99.5	mainlib	41.571	1379.28	26857595
Cyclohexane, 1-methylene-4-(1-methylethenyl)-	98.4	mainlib	19.175	1002.47	24896190
4-Hexen-3-one	83.2	replib	20.584	1028.53	22865417.35
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	21011932.5
Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	91.6	replib	18.3	984.45	18475592.98
Heptanoic acid, methyl ester	98.2	replib	19.987	1017.49	18384367.5
o-Isopropenyltoluene	97.6	mainlib	23.794	1087.95	18268530
Limonene	95.6	replib	20.544	1027.8	14994065

Tabla 85. Continuación Tabla 84...

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Butanoic acid, 3-methyl-, heptyl ester	87.4	mainlib	38.702	1333.26	14017877.5
Eucalyptol	99.6	mainlib	20.767	1031.91	13684362.5
6-Methylhept-4-en-1-yl 2-methylbutanoate	99.2	mainlib	41.179	1372.99	11903137.5
4-Methylpentyl 4-methylpentanoate	97.7	mainlib	37.122	1307.91	11769397.5
Propanoic acid, 2-methyl-, 4-methylpentyl ester	99.8	mainlib	24.755	1105.21	11287665
Pentanoic acid, cyclohexyl ester	89.2	replib	32.216	1229.17	11069067.5
6-Methylhept-4-en-1-yl isobutyrate	99.7	mainlib	35.648	1284.25	10985792.25
Phenethyl alcohol	94.5	flavor	8.783	763.4	10538175.25
Isoamyl isovalerate	97.9	flavor	24.481	1100.6	8722552
6-Methylheptyl 3-methylbutanoate	99.5	mainlib	42.557	1395.1	7897711
Styrene	99.9	mainlib	13.65	884.2	7300639
Methyl caproate; methyl hexanoate	99.4	flavor	14.944	914.08	7146177.5
Isoamyl hexanoate	91.4	flavor	30.978	1209.29	6647870.5
Formaldehyde oxime trimer	87.6	mainlib	18.58	990.3	6515793
(+)-2-Bornanone	99.9	replib	27.273	1147.5	6395974.75
1-Pentanol, 4-methyl-	96.8	replib	11.114	819.85	6237391.25
1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	97.9	replib	21.203	1039.99	6082063.75
Isovaleric acid	93.5	flavor	11.019	817.43	4624694.425
Butanoic acid, 2-methyl-, heptyl ester	99.8	replib	38.315	1327.05	4584959.75
7-Oxabicyclo[2.2.1]heptane, 1-methyl-4-(1-methylethyl)-	98.7	mainlib	19.674	1011.7	4511512.75

Tabla 86. Áreas relativas de metabolitos volátiles de la accesión Miscucho, (ají liofilizado)

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Benzaldehyde	96.8	replib	17.052	958.28	131308675
Butyl butyrate	97.7	flavor	18.552	989.72	78979340.5
2(3H)-Furanone, 5-ethylidihydro-	85.2	mainlib	30.414	1200.24	70568685
Terpinolene	93.4	flavor	23.571	1083.81	57024390
o-Cymene	92.9	mainlib	20.245	1022.26	50686570
Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	91.6	replib	18.3	984.45	48050667.5
4-Methylpentyl 2-methylbutanoate	99.5	mainlib	30.02	1193.64	36248227.5
Cyclohexane, 1-methylene-4-(1-methylethenyl)-	98.4	mainlib	19.175	1002.47	32121612.5
4-Hexen-3-one	83.2	replib	20.584	1028.53	29598187.5
6-Methylhept-4-en-1-yl 3-methylbutanoate	99.5	mainlib	41.571	1379.28	29534620
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	25484875
Benzene, 2-ethyl-1,4-dimethyl-	86.2	mainlib	19.928	1016.39	25181207.35
Sebacic acid, di(2-chlorophenethyl) ester	100	mainlib	21.971	1054.2	22560994.34
Pyrazine, tetramethyl-	86.2	replib	23.407	1080.78	20667312.95
o-Isopropenyltoluene	97.6	mainlib	23.794	1087.95	20317535
Heptanoic acid, methyl ester	98.2	replib	19.987	1017.49	19858700
Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	99.5	mainlib	15.228	920.04	19711995
beta-Pinene	99.7	flavor	17.877	975.58	18054220
gamma-Terpinene	99.6	null	21.986	1054.48	16703210
delta-3-carene	98.1	flavor	15.649	928.87	16184485

Tabla 87. Continuación Tabla 86...

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Butanoic acid, 3-methyl-, heptyl ester	87.4	mainlib	38.702	1333.26	15480695
alpha-terpinene	99.6	flavor	19.8	1014.03	15391650
4-Methylpentyl 4-methylpentanoate	97.7	mainlib	37.122	1307.91	14369215
Propanoic acid, 2-methyl-, 4-methylpentyl ester	99.8	mainlib	24.755	1105.21	14364092.5
Pentanoic acid, cyclohexyl ester	89.2	replib	32.216	1229.17	14240370
6-Methylhept-4-en-1-yl isobutyrate	99.7	mainlib	35.648	1284.25	13499665
6-Methylhept-4-en-1-yl 2-methylbutanoate	99.2	mainlib	41.179	1372.99	13319360
4-Decen-6-yne, (Z)-	84.5	mainlib	17.58	969.35	11987448.25
Methyl caproate; methyl hexanoate	99.4	flavor	14.944	914.08	11684802.5
Isoamyl isovalerate	97.9	flavor	24.481	1100.6	11614617.5
Triacetin	99.9	flavor	38.588	1331.43	9097161.25
Isoamyl hexanoate	91.4	flavor	30.978	1209.29	8302744.25
Phenethyl alcohol	94.5	flavor	8.783	763.4	7787148
6-Methylheptyl 3-methylbutanoate	99.5	mainlib	42.557	1395.1	7660325.25
Formaldehyde oxime trimer	87.6	mainlib	18.58	990.3	6497697.75
Eucalyptol	99.6	mainlib	20.767	1031.91	6148052
1-Pentanol, 4-methyl-	96.8	replib	11.114	819.85	6145976.5
Styrene	99.9	mainlib	13.65	884.2	6019413.75
1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	97.9	replib	21.203	1039.99	5679151.5
Ethyl butyrate	99.8	flavor	9.738	785.96	5229837.25

Tabla 88. Áreas relativas de metabolitos volátiles de la accesión Tomatito rojo, (ají liofilizado)

Nombre del metabolito	Dot product	dbs	t _R (min)	índice de Kovats	promedio de áreas
Benzaldehyde	96.8	replib	17.052	958.28	134602750
Butyl butyrate	97.7	flavor	18.552	989.72	78114765
Terpinolene	93.4	flavor	23.571	1083.81	71039007.5
o-Cymene	92.9	mainlib	20.245	1022.26	51315472.5
2(3H)-Furanone, 5-ethylidihydro-	85.2	mainlib	30.414	1200.24	49686350
Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	91.6	replib	18.3	984.45	36259077.5
Cyclohexane, 1-methylene-4-(1-methylethenyl)-	98.4	mainlib	19.175	1002.47	32199937.5
Sebacic acid, di(2-chlorophenethyl) ester	100	mainlib	21.971	1054.2	31042367.5
o-Isopropenyltoluene	97.6	mainlib	23.794	1087.95	26628442.5
Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	99.5	mainlib	15.228	920.04	22282167.5
beta-Pinene	99.7	flavor	17.877	975.58	20209357.5
delta-3-carene	98.1	flavor	15.649	928.87	19885180
gamma-Terpinene	99.6	null	21.986	1054.48	17855752.5
Heptanoic acid, methyl ester	98.2	replib	19.987	1017.49	17005072.5
4-Methylpentyl 2-methylbutanoate	99.5	mainlib	30.02	1193.64	16460015
alpha-terpinene	99.6	flavor	19.8	1014.03	15706495
4-Decen-6-yne, (Z)-	84.5	mainlib	17.58	969.35	15036250
Isoamyl hexanoate	91.4	flavor	30.978	1209.29	14628147.5
Limonene	95.6	replib	20.544	1027.8	12057227.5
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	11334216.25

Tabla 89. Continuación Tabla 88...

Nombre del metabolito	Dot product	dfs	tr (min)	índice de Kovats	promedio de áreas
Butanoic acid, 3-methyl-, hexyl ester	99.8	replib	32.596	1235.27	11334216.25
4-Hexen-3-one	83.2	replib	20.584	1028.53	11240764.35
Isoamyl isovalerate	97.9	flavor	24.481	1100.6	10645557
4-Methylpentyl 4-methylpentanoate	97.7	mainlib	37.122	1307.91	9560612.5
Propanoic acid, 2-methyl-, 4-methylpentyl ester	99.8	mainlib	24.755	1105.21	7725325.5
Phenethyl alcohol	94.5	flavor	8.783	763.4	7640388.25
Pentanoic acid, cyclohexyl ester	89.2	replib	32.216	1229.17	7146265.75
Methyl caproate; methyl hexanoate	99.4	flavor	14.944	914.08	7074589.25
Isovaleric acid	93.5	flavor	11.019	817.43	6981493.25
Eucalyptol	99.6	mainlib	20.767	1031.91	5935597
Ethyl butyrate	99.8	flavor	9.738	785.96	5914416.75
Styrene	99.9	mainlib	13.65	884.2	5808091.5
6-Methylhept-4-en-1-yl 3-methylbutanoate	99.5	mainlib	41.571	1379.28	5739690.75
1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	97.9	replib	21.203	1039.99	5510262.25
Hexyl hexanoate	97.1	flavor	39.372	1344.01	5357826
Hexanoic acid, 4-hexen-1-yl ester	94.8	mainlib	39.049	1338.82	4600671.5
Butanoic acid, 3-methyl-, heptyl ester	87.4	mainlib	38.702	1333.26	4327034
7-Propylidene-bicyclo[4.1.0]heptane	92.7	mainlib	18.089	980.02	3950820.75
Hexanoic acid, ethyl ester	98.9	replib	18.663	992.05	3681717.25
2,6-Dimethyl-2-trans-6-octadiene	79.6	mainlib	18.728	993.42	3524747