Food Chemistry 386 (2022) 132704

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Metabolomic characterization of 5 native Peruvian chili peppers (*Capsicum* spp.) as a tool for species discrimination

Fabio Espichán [a,*], Rosario Rojas [a], Fredy Quispe [b], Guillaume Cabanac [c], Guillaume Marti [d,e]

[a] Universidad Peruana Cayetano Heredia (UPCH), Lima, Peru
[b] Instituto Nacional de Innovación Agraria (INIA), Lima, Peru
[c] Institut de Recherche en Informatique de Toulouse, Université de Toulouse, UT3, CNRS, Toulouse, 31400, France
[d] UMR 152 Pharmadev, Université de Toulouse, IRD, UT3, France
[e] Metatoul-AgromiX Platform, LRSV, Université de Toulouse, CNRS, UT3, INP, Toulouse, France

ARTICLE INFO

Keywords:
Metabolomic
Capsicum
OPLS-DA
SUS-plot
Biomarkers
Peruvian native chili peppers

ABSTRACT

Many species of chili peppers have overlapping morphological characters and delimitation by visual descriptors in many cases fails to differentiate one species from another. In Peru, there are 413 accessions of native chili pepper and 296 accessions of rocotos conserved in the Germplasm Collections of the National Institute of Agrarian Innovation (INIA), of which five accessions (three species from three locations) were selected for the present metabolomic study. The Discrimination of the three species of native chili peppers and identification of biomarkers was performed using untargeted metabolomic approach based on profiling by UHPLC-HRMS and multivariate data analysis. The samples of fresh chili peppers (whole fruit) from Chincha area were used to construct an OPLS-DA model. To validate the biomarkers (identified 15 biomarkers, mainly flavonoids), an external validation set of the OPLS-DA model was constructed using Chiclayo and Huaral collection datasets. Consequently, the OPLS-DA based on Chincha samples model has a high predictive capacity demonstrating that the biomarkers have a high probability of continuity in any culture space, being successful in discriminating the species by untargeted metabolomics.

1. Introduction

The Capsicum genus (Solanaceae) exhibits a remarkable diversity of aromas, colors, flavors, fruit shapes and growth habits. From the point of view of classical taxonomy, many species also casually present similar morphological characters that fail to differentiate one species from the other (Walsh and Hoot 2001). This problem is evident, for example, in the delimitation of the species *Capsicum annuum*, *C. chinense* and *C. frutescens*, which form a taxonomic complex poorly differentiated (Hernández-Verdugo, Dávila, and Oyama, 2017). The *Capsicum annuum* species can be difficult to separate from the point of view of classical taxonomy from cultivated *C. chinense* (the hottest pepper) and *C. frutescens* (Tabasco pepper) and their morphological characteristics can overlap. These three species share the same ancestral gene pool and are sometimes referred to as the "annuum-chinense-frutescens complex" (Zhigila et al., 2014). Despite the limitations in the classification and identification of species in the genus *Capsicum*, 38 species have been identified, of which *C. annuum*, *C. chinense*, *C. frutescens*, *C. baccatum* and *C. pubescens* are found as the five species domesticated in Peru and the world (Eshbaugh, 1983; Bosland and Votava, 2012).

At present, identification studies of the species from genus *Capsicum* are carried out with molecular tools such as AFLP (amplified fragment length polymorphism) and cytogenetics (Moscone et al., 2014; Patricia Toquica et al., 2003; Wahyuni et al., 2013). Metabolomics has the advantage of studying the chemical phenotype in comparison with molecular tools that investigate genetic information. The phenotype refers to the genotypic expression of an organism and is influenced by the environmental characteristics in which the genes are expressed. Due to this environmental influence, organisms express different phenotypes as they develop (Orgogozo, Morizot, and Martin, 2015). Metabolomic studies can also provide metabolomic phenotype information that can be used as a guiding strategy for genetic improvement programs for chili peppers with specific phenotypic quality traits (Wahyuni et al., 2013).

High-performance liquid chromatography coupled with high-

* Corresponding author.
E-mail addresses: fabio.espichan.j@upch.pe (F. Espichán), Rosario.rojas@upch.pe (R. Rojas), equispe@inia.gob.pe (F. Quispe), guillaume.cabanac@univ-tlse3.fr (G. Cabanac), guillaume.marti@univ-tlse3.fr (G. Marti).

https://doi.org/10.1016/j.foodchem.2022.132704
Received 1 October 2021; Received in revised form 7 March 2022; Accepted 12 March 2022
Available online 17 March 2022

resolution mass spectrometry is currently the most widely used approach for studying metabolic phenotypes through targeted and untargeted analyses (H. Gika et al., 2019). Therefore, the untargeted metabolomic fingerprints are proposed as an alternative taxonomic descriptor for species discrimination through biomarker identification for the characterization of these native species.

The importance of studying chili peppers (*Capsicum spp*.) lies in that is one of the most important crops in the world (Ortiz et al., 2010); and its fruits express important sources of metabolites beneficial to human health such as ascorbic acid (vitamins C), tocopherols (Vitamin E), carotenoids (provitamin A), flavonoids (antioxidants) and capsaicinoids, the latter being responsible for its pungency (Wildman, 2007).

Currently, efforts are being made in Peru to rescue and conserve native chili peppers. Thus, in 2013, the National Institute of Agrarian Innovation (INIA), with the collaboration of Bioversity International (Libreros et al., 2013) managed to collect and make possible the conservation of 413 accessions of chili peppers of the species *C. annuum*, *C. baccatum*, *C. chinense* and *C. frutescens* and 299 accessions of *C. pubescens* from 11 departments of Peru. The conservation of these collected chili peppers was carried out at the INIA Donoso Agrarian Experimental Station located in the Huaral province of the Lima region, the assignment of the species to the conserved materials (accessions) was determined by comparing their taxonomy and agro-morphological characteristics of the fruit, flowers, leaves and other phenotypic traits. Given the great diversity of native peppers, the metabolomic study and its relationship with the assigned species are pending (Libreros et al., 2013; Meckelmann et al., 2013). In 2016, INIA in collaboration with the Universidad Peruana Cayetano Heredia (UPCH) carried out the agro-morphological, chemical, nutritional, antioxidant and sensory characterization of 32 accessions preserved by INIA, 18 materials from Chincha, Pucallpa and Chanchamayo (Patel et al., 2016). From this last study, five promissory materials were identified based on their agronomic behavior and recommendations of gourmets (Chefs). The present metabolomic study aims to the characterization and identification of biomarkers for the discrimination of these three species from the Peruvian diversity.

In order to establish the link between species and chemical fingerprints, samples of five accessions (charapita, dulce rojo, miscucho, ayucllo and tomatito rojo) from three species (*C. baccatum*, *C. frutescens*, *C. chinense*) harvested in three distinct areas (Chincha, Chiclayo and Huaral) profiled by LC-MS. These five native accessions of the present study, have not been domesticated and have not carried out molecular studies or non-targeted metabolomics. At present, only a characterization of species has been carried out by classical taxonomy of morphological similarity: *C. frutescens* (charapita), *C. baccatum* (ayucllo and tomatito rojo), *C. chinense* (miscucho and dulce rojo) (see Table 1).

2. Material and methods

2.1. Plant material

The conservation of native chili pepper was made at the National Collection of germplasm at the Donoso Experimental Station (Central Peruvian coast) of INIA. The five accessions were previously taxonomically characterized by classical visual descriptors by the INIA and grouped into three species. The ayucllo (*C. baccatum*) and Charapita (*C. frutescens*) accessions were originally collected in the San Martin region of Peru (Central forest). Also, the tomatito rojo (*C. baccatum*) and dulce rojo (*C. chinense*) accessions were originally collected in the Ucayali region (Central forest). The Miscucho accession (*C. chinense*) was originally collected in the Lambayeque region (North Coast). The five accessions of chili peppers were transferred from the National Collection of germplasm to the experimental places at the Peruvian coast.

The native chili peppers were cultivated between September and December 2016 in three locations on the Peruvian coast and harvested in January and March 2017. The places were the following: 1) Nursery of the Agroexport Topara Company located in Chincha Alta 200 km south of Lima, 13° 12′32.33″ south latitude, west longitude 76° 9′22.78″ and altitude of 448 masl, temperatures between 18 and 20 °C and relative humidity 77–78%; 2) Vista Florida Experimental Station from the district of Picsi, province of Chiclayo, department of Lambayeque with the south latitude of 06° 43′34″, west longitude 79° 46′49″ at the altitude of 30 masl with average temperature 30° C, 800 mm precipitation and 70% relative humidity; 3) Donoso Experimental Station located in Huaral department of Lima (Fig. 1) at 11° 31′22.8″ south latitude, west longitude 77° 13′53.6″ and altitude of 180 masl, temperatures between 27.8 and 28.8 °C and relative humidity between 79% −80%. The production of chili peppers from Chincha was organic agricultural management, while for the other two localities the agricultural management was conventional. A representative sample of each native chili accession (between 5 and 10 kg) was collected from each of the three harvest locations: Chincha, Chiclayo, and Huaral. From each representative sample, 50 g of fresh whole native peppers per accession were lyophilized. Freeze-dried native peppers were ground and stored at −20 °C for later analysis by UHPLC-HRMS.

2.2. Metabolite extraction and profiling UHPLC-HRMS

The extraction of the metabolites of the 15 lyophilized chili peppers (Table 1) from the three locations (Fig. 2A, 2B) was carried out in triplicate with 80% methanol (HPLC grade, Thermo fisher/water v/v) (type I water, 18 MΩ.cm), according to the following steps: 1) In eppendorf centrifuge microvials of 1.7 μL, 10 mg of lyophilized chili

Table 1
Total samples of peppers by harvest place and analytical samples.

Number of samples	Accession	Species	Analytical replicates	Harvest place	Agricultural management	Sample set for external validation of biomarqueurs
1	Ayucllo	*C.baccatum*	4	Chincha	Organic	*Training set*
2	Tomatito rojo	*C.baccatum*	4			
3	Dulce rojo	*C.chinense*	4			
4	Miscucho	*C.chinense*	4			
5	Charapita	*C.frutescens*	4			
6	Ayucllo	*C.baccatum*	4	Chiclayo	Conventional	*Test set*
7	Tomatito rojo	*C.baccatum*	4			
8	Dulce rojo	*C.chinense*	4			
9	Miscucho	*C.chinense*	4			
10	Charapita	*C.frutescens*	4			
11	Ayucllo	*C.baccatum*	4	Huaral	Conventional	*Test set*
12	Tomatito rojo	*C.baccatum*	4			
13	Dulce rojo	*C.chinense*	4			
14	Miscucho	*C.chinense*	4			
15	Charapita	*C.frutescens*	4			

Table 2
Results of the SUS-plot analysis and identification of the biomarkers of the species *C. baccatum*, *C. chinense* and *C. frutescens* of the chili peppers of Chincha location.

Species	RT (min)	Exact mass	Molecular formula	Error (Da)	Adduct	Metabolite name	Chemical class	Source data base[a]	Final score[b]
C. baccatum	3.805	610.132	$C_{30}H_{26}O_{14}$	−0.0002209	$[M-H]^-$	3′,4′,5,7,8-Pentahydroxyflavone; 7-O-[4-Hydroxy-E-cinnamoyl-(→6)-β-D-glucopyranoside]	Flavonoid	*Capsicum*	15.236
C. baccatum	2.841	350.064	$C_{16}H_{14}O_9$	0.0083056	$[M-H]^-$	3,5,7-trihydroxy-6-methoxy-2-(2,4,5-trihydroxyphenyl) −2,3-dihydro-1-benzopyran-4-one	Flavonoid	*Capsicum*	14.3606
C. baccatum	4.344	1080.535	$C_{51}H_{84}O_{24}$	0.00953	$[M+H]+$	Spirostane-2,3-diol; 3-O-[β-D-Galactopyranosyl-(1 → 2)-[b-D-glucopyranosyl-(1 → 3)]-β-D-glucopyranosyl-(1 → 4)-β-D-galactopyranoside]	Steroid	*Capsicum*	11.4656
C. baccatum	3.908	581.151	$C_{26}H_{29}O_{15}$	0.0050966	$[M]+$	Cyanidin 3,5-diglycosides; 3-O-β-D-Xylopyranoside, 5-O-β-D-glucopyranoside	Anthocyanin	*Capsicum*	15.519
C. baccatum	3.845	568.194	$C_{30}H_{32}O_{11}$	−0.0016146	$[M-H]^-$	Hypochoeroside D;(+) -Hypochoeroside D	Cinnamic acids and derivatives	UNPD	7.5549
C. frutescens	4.628	572.226	$C_{30}H_{36}O_{11}$	−0.0044145	$[M-H]^-$	Physalin B; 2,3-Dihydro, 2α-ethoxy, 5β,6β-epoxide	Steroid	Solanaceae	9.8859
C. frutescens	3.857	432.088	$C_{21}H_{20}O_8S$	−0.0026879	$[M-H]^-$	Roseochelin B	Carboxylic acid	Npatlas	7.6464
C. frutescens	2.559	358.069	$C_{18}H_{14}O_8$	−0.009009	$[M-H]^-$	3,5-dihydroxy-2-(2-hydroxyphenyl)-7-methoxy-4-oxochromen-8-yl acetate	Flavonoid	*Capsicum*	15.2364
C. frutescens	5.498	556.231	$C_{30}H_{36}O_{10}$	−0.0055291	$[M-H]^-$	Cerberalignan J; 8′-Deoxy	Lignan	*Capsicum*	13.4496
C. frutescens	4.944	868.368	$C_{50}H_{52}N_4O_{10}$	0.0045674	$[M-H]^-$	*N*-acetyl-D-glucosaminyl-(1 → 4)-*N*-acetylmuramoyl-L-alanyl-D-glutamyl-6-carboxy-L-lysine	Aminoacids	ChEBI	6.1943
C. chinense	1.049	328.095	$C_{18}H_{16}O_6$	0.0052118	$[M-H]^-$	2′,6,8-Trihydroxy-7-(hydroxymethyl) flavone; 2′,6-Di-Me ether	Flavonoid	*Capsicum*	15.494
C. chinense	2.847	444.178	$C_{24}H_{28}O_8$	−0.0058586	$[M-H]^-$	3,4′,5,6,7-Pentahydroxy-3′-(8-hydroxy-7-methyloctyl) flavone	Flavonoid	*Capsicum*	14.202
C. chinense	10.451	710.388	$C_{37}H_{58}O_{13}$	0.0037655	$[M-H]^-$	Mandelalide B	Macrolide	UNPD	6.6717
C. chinense	0.74	224.053	$C_7H_{12}O_8$	0.0077409	$[M-H]^-$	2,4,5-trihydroxy-3-methoxy-1,6-hexanedioic acid	Carboxylic acid	UNPD	7.1142
C. chinense	10.547	710.388	$C_{37}H_{58}O_{13}$	0.0039655	$[M-H]^-$	2,3,16,23-Tetrahydroxy-12-oleanene-24,28-dioic acid; 2-Me ether, 3-O-β-D-glucopyranoside	Terpenoid	*Capsicum*	12.7498

[a]**Source: family** = Solanaceae, genus = Capsicum, (from DNP = Dictionary of natural products), UNPD = Universal Natural Products Database, Npatlas = Natural products Atlas.
[b]**Final scores** = from MSClean input multiplier of: genus×2, family ×1.5, generic ×1.

pepper were weighed in triplicate, 800 μL of 80% methanol was added, then 1 min by vortex and 10 min by ultrasound and again 1 min by vortex, 2) Samples were centrifuged using an eppendorf model 5804 R centrifuge at 14,000 rpm/20913 *xg* for 10 min at 4° C; 3) then 600 μL of the supernatant was taken and transferred to another centrifuge microvial, 4) Steps 1 to 3 were repeated for the second extraction, obtaining a total volume of 1200 μL of extract, this was done in triplicate 5) From each sample group (from the triplicate) a new sample was prepared consisting of 200 μL of each, which was considered as a "fourth replica" whose final volume was 600 μL. Also, samples were prepared for blank and quality control (QC). The blank consisted of a sample of 80% aqueous methanol solvent, subjected to steps 1–4 of the previous extraction procedure, and the QC was prepared by taking 100 μL of each "fourth replica", until a volume of 1500 μL was obtained. The extracts of the 15 × 4 repeats of native peppers were injected into the UHPLC system. The QCs and blanks were injected into the UHPLC system intermittently every 3 samples, resulting in a total of 6 QCs samples and six blanks. The samples were profiled in the two ESI ionization modes (+, −) using the UHPLC – DAD – LTQ Orbitrap XL instrument (Ultimate 3000, Thermo Fisher Scientific, Hemel Hempstead, UK). In UHPLC, UV detection was performed using a 210–400 nm diode array detector; Runs were performed in binary gradient mode using a Waters Acquity UHPLC BEH C18 Particle size 1.7 μm, 2.1x100 mm column, ULTRA Holder column guard (Phenomenex UHPLC 2.1 at 4.6 mm ID). Mobile phase A: H_2O + 0.1% FA; and phase B: ACN + 0.1% FA; (FA = formic acid, LC-MS grade, Sigma Aldrich, ACN = acetonitrile, HPLC grade, Thermo fisher) the gradient conditions were: 0 → 5 min 95% A, 5% B; 5 → 12.2 min, 5% A, 95% B (Chervin et al., 2017). In HRMS Orbitrap, mass detection was performed using an ESI (+, −) electrospray source with a resolving power of 15,000 (full width at half maximum (FWHM) at 400 m / z). The mass scan range was 100–2000 *m/z*. The capillary temperature was 300 °C and the spray voltage was set at 4 kV (positive mode) and 4.2 kV (negative mode). The mass measure was externally calibrated before starting the experiment. Each full mass was followed by MS/MS data-dependent on the three most intense peaks using collision-induced dissociation (CID) stages (35% normalized collision energy, isolated with 2 Da width, Q 0.250 activation). At the end of the analysis, a total of 72 chromatograms (60 extracts, 6 QCs, 6 Blanks) were obtained (Fig. 2 C). The chromatograms of the QC samples in both ionization modes are shown in Figs. S1 and S2 (supplementary information).

2.3. Data processing

The conversion of the raw data (72 chromatograms) from Thermo *. RAW format to *.ABF format was carried out with ABF Convert software (Reifycs Abf Converter, https://www.reifycs.com/AbfConverter/) and processed with MS-DIAL version 4.48 (Tsugawa et al., 2015). For the alignment and deconvolution of the characteristics or metabolites, it was set at a *m/z* range of 100 to 2000 Da and retention time of 0 and 15 min. The tolerance MS1 and MS2 were established at 0.01 and 0.4 Da respectively in centroid mode. The minimum peak height or Amplitude (A) was calculated by MS-DIAL experiments by interpolation of the plot Amplitude versus the number of MS2 features in each ionization mode. For ESI (+) a number of features of 189 (A = 15E5) and 193 for ESI (−) (A = 10E5) were approximated for each harvest of peaks. The peaks were aligned on a quality control (QC) file with a mass tolerance of 0.015 Da and a retention time tolerance of 0.05. The aligned results were exported in *.TXT format (Raw data matrix and normalized data matrix)

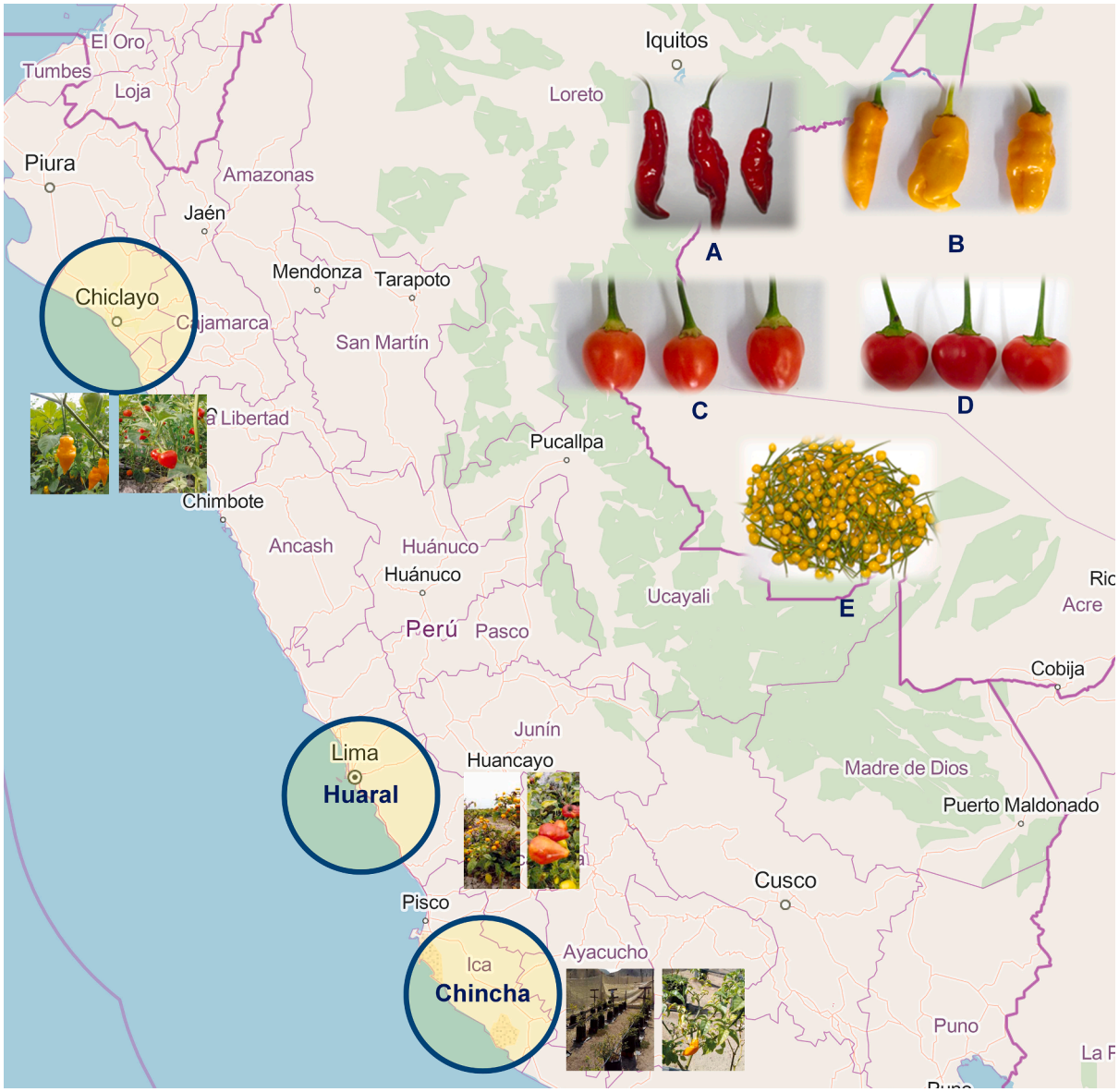

Fig. 1. Map of Peru showing the three zones of the Peruvian coast Chincha, Chiclayo and Huaral where the Accessions of Native Chilies are harvested: A) Dulce Rojo (*C. chinense*), B) Miscucho (*C. chinense*), C) Ayuclo (*C. baccatum*), D) Tomatito Rojo (*C. baccatum*), E) Charapita (*C. frutescens*). Map obtained from https://www.arcgis.com.

and the total number of "all peaks" of each ionization mode for data treatment using the MS-CleanR workflow (Fraisier-Vannier et al., 2020).

2.4. Cleaning the data and annotation of metabolites

The MS-CleanR package was used to tackle degenerate features and filter artifacts and noise signals. The filters were selected using a default value for cleaning the MS-DIAL peak list table: with a minimum blank ratio set to 0.8, a maximum relative standard deviation (RSD) set to 30, and a relative mass defect (RMD) ranging from 50 to 3000. The maximum mass difference for the detection of features relationships was established at 0.005 Da and the maximum retention time (RT) difference was established at 0.025 min. The Pearson correlation to compute cluster was set with correlation ≥ 0.8 and statistically significant $\alpha = 0.05$. Then MS-CleanR uses the MSCombine (Calderón-Santiago et al., 2016) tool to combine the ESI data tables (+, −) and to maintain positive and negative common features. After filtering, for ESI (+, −) the number of features was reduced to 228 and 269 respectively out of a total of 662 for ESI (+) and 783 for ESI (−). The filtered peaks were exported to MS-FINDER (version 3.5) (Lai et al., 2018). The annotation of metabolites was carried out with the MS/MS spectra mirrored to *in silico* spectra modeled by MS-FINDER using the generic database: KNApSAcK (Natural products), NANPDB (Natural products), PlantCyc (Plant), UNPD (Natural products), ChEBI (Biomolecules), DrugBank (Drug), Pubchem (Biomolecules), FooDB (FooD), STOOF (Environment), NPA (Natural products Atlas), COCONUT (Natural product) and LipidMAPS (Lipids), and the Dictionary of Natural Products (DNP on DVD, v26.2 CRC Press) database comprising compounds from the genus *Capsicum* and the family of Solanaceae. The annotation was prioritized according to the biological source levels for genus, family and generic (DNP and internal MS-FINDER database) respectively. The score was calculated as the value of the experimental fragmentation of MS2 reflected to *in silico* similarity. The total score was calculated using a multiplier associated with the biological source to prioritize the annotations. The multipliers 2: 1.5: 1 were considered for genus, family and generic respectively. The generated file contained the combined ESI (+, −) feature prioritization database with the MS-FINDER score levels was used for multivariate statistical analysis.

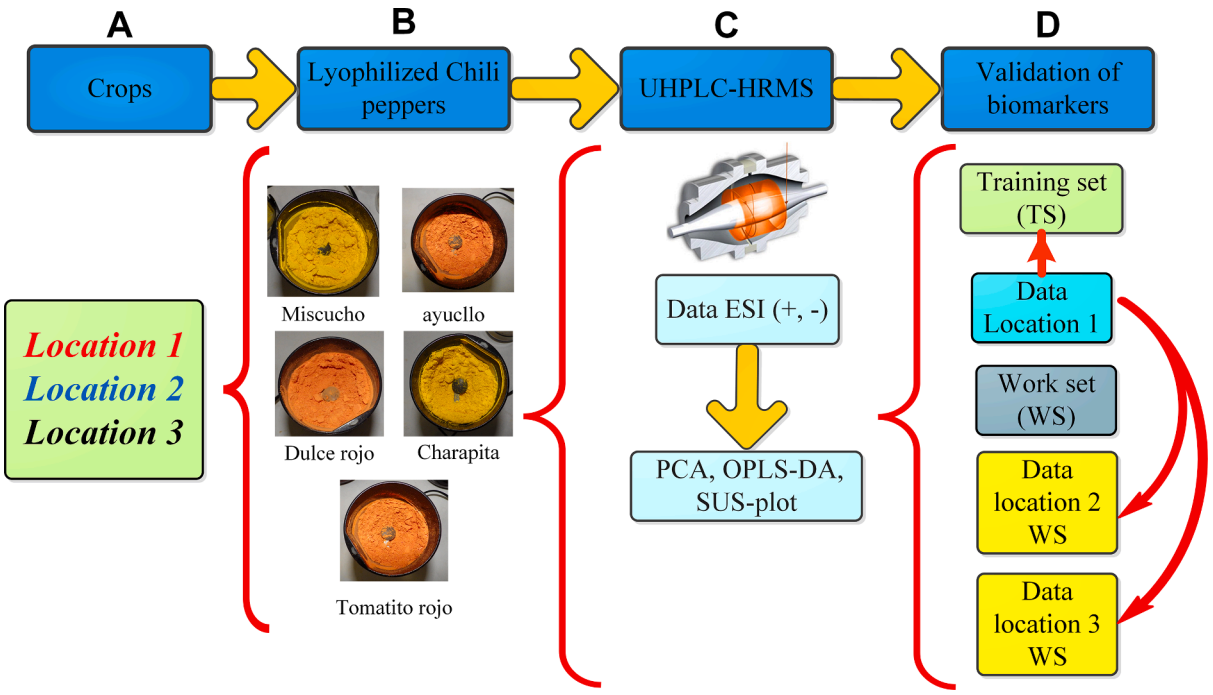

Fig. 2. Metabolomic workflow diagram by UHPLC-HRMS for the identification and validation of biomarkers.

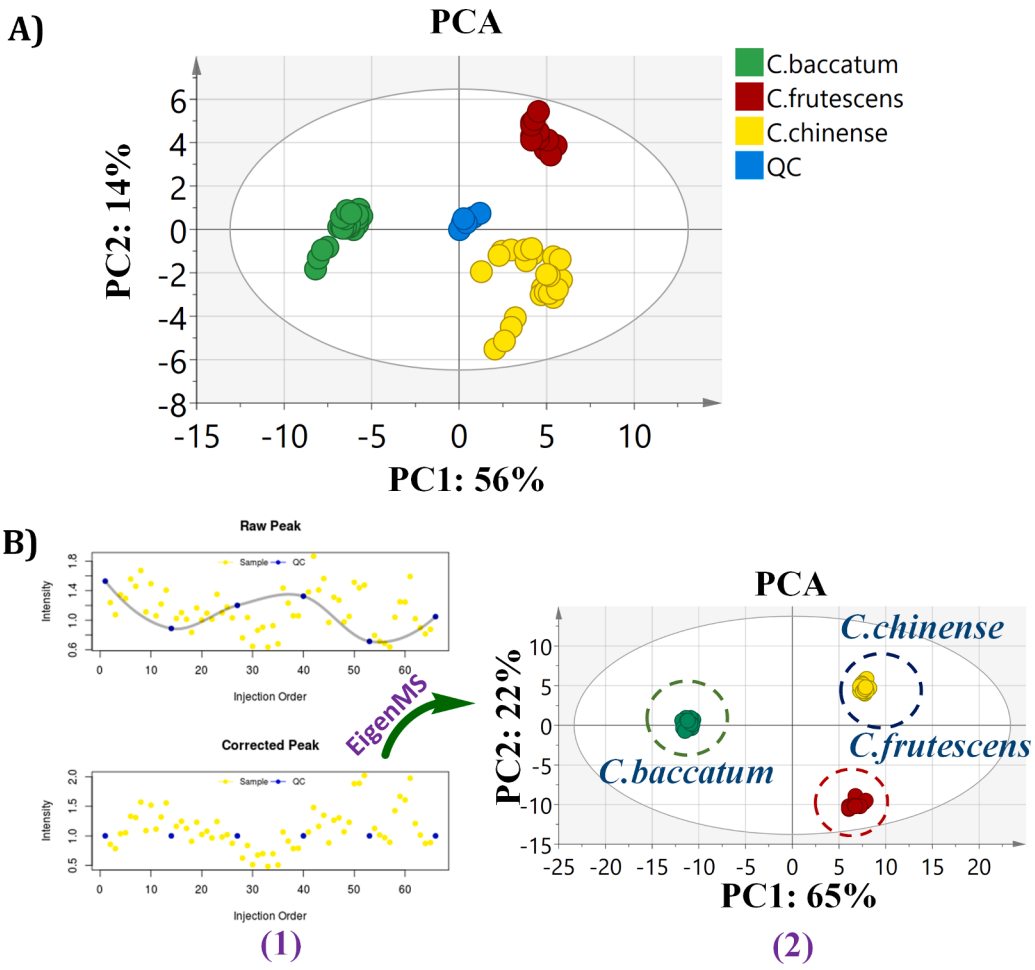

Fig. 3. The top (A) is of PCA of class species for samples of chili peppers (3 species) of three locations, Chincha, Chiclayo and Huaral, ($R^2 = 0.976$), by MS-CleanR approach, in B, 1) QC-based Correction of final data of MS-CleanR by EigenMS Normalization, 2) PCA by species after of Correction.

2.5. Data normalization

Principal component analysis (PCA), of the data obtained by MS-CleanR is shown in Fig. 3A. The normalization was carried out to eliminate sources of systematic variation originating from features related to growth, climate or agroecological treatment. This variation is observed as intraclass variation in PCA within each species of chili pepper. The online software NOREVA 1.0 was used for normalization (Li et al., 2017). The software performs a correction based on the QC information using the QC-RLSC algorithm. For this, a preprocessing of the data was carried out with the default parameters: Filter criterion = 0.8, Bias-variance tradeoff = 0.75, Regression model = Local polynomial fits and Imputation algorithm = K-Nearest Neighbour (KNN). Among the 20 normalization methods tested, the EigenMS normalization method was chosen (Karpievitch et al., 2014) as it presented the best performance. Finally, the performance of the normalization method was evaluated by evaluating the PCA and ROC graph (supplementary information, Figs. S5-S6).

2.6. Multivariate statistical analysis

2.6.1. Species discrimination

For species discrimination, a supervised OPLS-DA analysis (discriminant analysis of orthogonal projections of latent structures) of the data normalized by EigenMS was performed using SIMCA-P (version 15.1, Sartorius Stedim Biotech, Umetrics, Umeå, Sweden), (Fig. 2 C). The clusters formed (three species) faced one by one to form 2 new OPLS-DA models (*C. baccatum* vs *C frutescens* and *C. chinense* vs *C. frutescens*). These two OPLS-DA models were used to perform a Shared and Unique Structure plot (SUS-plot) analysis. In addition, a hierarchical Clustering Heatmap diagram was built for the 15 discriminant metabolites obtained from the SUS-plot analysis using Distance Measure Euclidean and the Clustering Algorithm Ward of the PLS-DA model, in which the discrimination of the 3 species stands out, using the software Metaboanalyst 5.0 (Xia et al., 2015).

Same discriminant models were tentatively constructed for separation of accessions and localities.

2.6.2. Biomarker validation

The validation of the OPLS-DA models was done using SIMCA P (Fig. 2 D). The raw data (three locations) was divided into three data blocks of 20 individuals each, with the same number of variables (metabolites) by locations, Chincha, Chiclayo and Huaral. The OPLS-DA model of the Chincha data was used as a training set (TS) and the data of the chili peppers from the town of Chiclayo and Huaral was used as a work set. The probability of fit to Chincha's OPLD-DA model was assessed using a threshold greater than 0.5. The same strategy was also carried out with 1/3 of the samples to model the "accessions" and 2/3 of the samples as test samples and observe the results.

3. Results

An electrospray ionization (ESI) interface, in both positive and negative ion modes, was used for the detection of metabolomic profiles, since they are complementary. The results of the chromatograms in the ESI (+, −) ionization modes are shown in Figs. S1 and S2 (supplementary information). In ESI (+), two chromatographic peaks with greater relative peak area were identified, *cis*-Capsaicin and Dihydrocapsaicin (Table 2).

The major compounds class, found in the UHPLC-HRMS profile in the five native peppers are flavonoids (12%), followed by carboxylic acids (8%), steroids (7%), phenols (5%). In phenols, seven capsaicinoids have been identified (Nordihydrocapsaicin, *cis*-Capsaicin, Homocapsaicin II, Homodihydrocapsaicin, N - [(4-hydroxy-3-methoxyphenyl) methyl] octanimidic acid, *N*-Vanillylnonanamide and Dihydrocapsaicin) (Table 3). The following are Carbohydrates (4%), others (28%), unknown (36%), Fig. S11 (supplementary information). This reveals the large number of metabolites yet to be discovered. In ESI (−), near the beginning of the chromatogram, compounds belonging to the genus Capsicum and the Solanacea family were identified. With the chemical class Benzoic acids and derivatives, two compounds were identified: 2-diazoniobenzoate and anthranilate that were found in the genus Capsicum. Some Carbohydrates like (2R, 3R, 4S, 5R, 6R) −2- (hydroxymethyl) −6 - {[(2S, 3R, 4R, 5R, 6R) −4,5,6-trihydroxy-2- (hydroxymethyl) oxan-3-yl] oxy} oxane-3,4,5-triol, were found in the family Solanaceae and 2,3-dioxo-L-gulonate was found in the genus Capsicum. Carboxylic acids such as citric acid and Malic acid were found in the Solanaceae family, and the compounds (3S) −3-amino-3-carboxypropanoate, (−) - quinate, 4-Hydroxy-2-pyrrolidinecarboxylic acid; *N*-Me, (2S) −5-amino-2 - [(1-hydroxyethylidene) amino] pentanoic acid and tyrosin are found in the genus Capsicum (Table 4). The acidic compounds with the greatest relative peak area are citric acid and maleic acid (see Fig. S2, supplementary information).

The boxplot diagrams and ANOVA analysis of *cis*-Capsaicin and Dihydrocapsaicin (Fig. S3, supplementary information) of their normalized relative areas did not show significant differences for the three species. While in the ESI (−) chromatograms, two peaks with the greater relative area were identified; citric acid (citrate) which shows a significant difference for the *C. baccatum* species compared to the other two species and could be considered as a biomarker for this species, and a low concentration of malic acid was also observed, which could be a differentiating metabolite of this species with respect to the species

Table 3
Capsaicinoids identified in the 5 native peppers.

RT (min)	Exact mass	Molecular formula	Error (Da)	Adduct	Metabolite name	Chemical class	Source data base[a]	Final score[b]
72.68	294.206	$C_{17}H_{27}NO_3$	0.0000702	[M + H]+	Nordihydrocapsaicin	Pseudoalkaloids	UNPD	8.3223
7.281	306.206	$C_{18}H_{27}NO_3$	−0.0000298	[M + H]+	Cis-Capsaicin	Alkaloids	*Capsicum*	13.479
7.807	320.222	$C_{19}H_{29}NO_3$	−0.0000798	[M + H]+	Homocapsaicin II	Alkaloids	*Capsicum*	15.073
8.249	322.238	$C_{19}H_{31}NO_3$	0.0001703	[M + H]+	Homodihydrocapsaicin	Alkaloids	*Capsicum*	13.656
7.727	308.221	$C_{18}H_{29}NO_3$	0.0008202	[M + H]+	Dihydrocapsaicin	Alkaloids	*Capsicum*	14.309
6.7	280.188	$C_{16}H_{25}NO_3$	0.0026201	[M + H]+	N-[(4-hydroxy-3-methoxyphenyl) methyl] octanimidic acid	Alkaloids	*Capsicum*	14.128
7.183	294.206	$C_{17}H_{27}NO_3$	0.0002702	[M + H]+	*N*-Vanillylnonanamide	Alkaloids	*Capsicum*	14.327

[a]**Source: genus** = Capsicum, (from DNP = Dictionary of natural products), UNPD = Universal Natural Products Database.
[b]**Final scores** = from MS-CleanR input multiplier of: genus ×2, family ×1.5, generic ×1.

Table 4
Carbohydrates and Carboxylic acids of 5 accessions of chili peppers.

Species	RT (min)	Exact mass	Molecular formula	Error (Da)	Adduct	Metabolite name	Chemical class	Source data base[a]	Final score[b]
C. baccatum	0.742	191.013	$C_6H_8O_7$	−0.0002209	$[M-H]^-$	Citrate	Carboxylic acid	Solanaceae	11.802
C. baccatum	0.739	133.009	$C_4H_6O_5$	0.0083056	$[M-H]^-$	Malic acid	Carboxylic acid	Solanaceae	12.336
C. baccatum	1.064	147.025	$C_7H_4N_2O_2$	0.00953	$[M-H]^-$	2-diazoniobenzoate	Benzoic acids and derivatives	*Capsicum*	12.151
C. baccatum	0.706	138.054	$C_7H_7NO_2$	0.0050966	$[M+H]^+$	Anthranilate	Benzoic acids and derivatives	*Capsicum*	12.903
C. frutescens	0.676	132.026	$C_4H_7NO_4$	−0.0026879	$[M-H]^-$	(3S)-3-amino-3-carboxypropanoate	Carboxylic acid	*Capsicum*	14.244
C. frutescens	0.702	191.049	$C_7H_{12}O_6$	−0.009009	$[M-H]^-$	(−)-quinate	Carboxylic acid	*Capsicum*	15.318
C. frutescens	0.709	146.079	$C_6H_{11}NO_3$	−0.0055291	$[M+H]^+$	4-Hydroxy-2-pyrrolidinecarboxylic acid; *N*-Me	Carboxylic acid	*Capsicum*	15.723
C. frutescens	0.726	175.106	$C_7H_{14}N_2O_3$	0.0045674	$[M+H]^+$	(2S)-5-amino-2-[(1-hydroxyethylidene) amino]pentanoic acid	Carboxylic acid	*Capsicum*	15.315
C. chinense	0.896	182.079	$C_9H_{11}NO_3$	0.0052118	$[M+H]^+$	Tyrosin	Carboxylic acid	*Capsicum*	15.409
C. baccatum	0.73	365.102	$C_{12}H_{22}O_{11}$	−0.0016146	$[M+Na]^+$	(2R,3R,4S,5R,6R)-2-(hydroxymethyl)-6-{[(2S,3R,4R,5R,6R)-4,5,6-trihydroxy-2-(hydroxymethyl)oxan-3-yl]oxy}oxane-3,4,5-triol	Carbohydrates	Solanaceae	11.547
C. frutescens	0.892	191.013	$C_6H_8O_7$	−0.0044145	$[M-H]^-$	2,3-dioxo-L-gulonate	Carbohydrates	*Capsicum*	15.589

[a]**Source: family** = Solanaceae, genus = Capsicum, (from DNP = Dictionary of natural products).
[b]**Final scores** = from MS-CleanR input multiplier of: genus ×2, family ×1.5, generic ×1.

C. frutescens and *C. chinense* (supplementary information, Fig. S3). In general, in the ESI (+) ionization mode for chili samples, polar compounds are observed at the beginning of the chromatogram that are common to sugars and organic acids. For ESI (−), metabolites with a greater area at the beginning of the chromatogram and semi-polar metabolites are observed in the center of the chromatogram. The alignment, deconvolution, and the use of the MS-CleanR package allowed the annotation of 162 metabolites out of a total of 182 metabolites in both ionization modes. The annotations (based on MS2 spectral similarity) were at significance level 2 according to the Metabolomic Standard Initiative (Sumner et al., 2007). First, a Principal Component Analysis was used to show patterns in few dimensions, as well as exploring the relationships between independent and dependent variables (Fig. 3A). The PCA score plot showed that independent variables such as agro-ecological practices and climatic conditions could modify the response (relative areas of metabolites). No outliers were observed since the samples were distributed within the Hotelling normality area of the ellipse with 95% confidence. It was possible to formulate as a first hypothesis three classes or clusters of samples differentiated by "species". The QCs (quality control) were plotted near the center of the PCA denoting a good overall reproducibility and robustness during LC-MS profiling of all samples (Broadhurst et al., 2018; H. G. Gika et al., 2007). Also, the effect of QC-based signal drift correction on intensities is observed (Fig. 3B-1). As an example, a selected feature is given by m/z = 151.03502 (To see other examples see supplementary information, Fig. 4). The correction was made using the QC-RLSC algorithm of the NOREVA Online platform. The QC-RLSC was used to estimate data quality and eliminate unwanted variations (features). This procedure was performed after the data imputation procedure (Luan et al., 2018). The variations observed could be of biological, agroecological or instrumental origin, such as the performance of the instrument, loss of signal intensity due to the run time, performance of the LC column, accumulation of contaminants in the source of MS ions or MS sensitivity, among others (Karpievitch et al., 2014). These variations were seen as an intraclass variation in PCA. To improve the PCA results, we proceeded with the normalization of the chromatographic data with dimensions [182x66] columns × rows. The EigenMS normalization method and the correction procedure based on the QC-RLSC allowed eliminating irrelevant information for the categorical classification by species. After normalization by species, it was observed that 13 "errors" of unknown origin were eliminated from the original table (metabolites), obtaining a final data matrix of [169x66]. To evaluate the performance of the normalization method and its ability to reduce intraclass variation, the PCA and ROC curves (receiver operating characteristic curves) were evaluated to show the balance between sensitivity and specificity (Alonso et al., 2015). The PCA results show good inter-species separation with an explained variance of 90% by species and AUC = 1 (supplementary information, Fig. S5-S6). The normalized table was exported to the SIMCA-P software to perform the multivariate OPLS-DA analysis using Pareto scaling. The OPLSDA was performed on Chincha samples and then evaluated for validity using other locations. Fig. 4A shows the OPLS-DA model for the discriminant analysis by

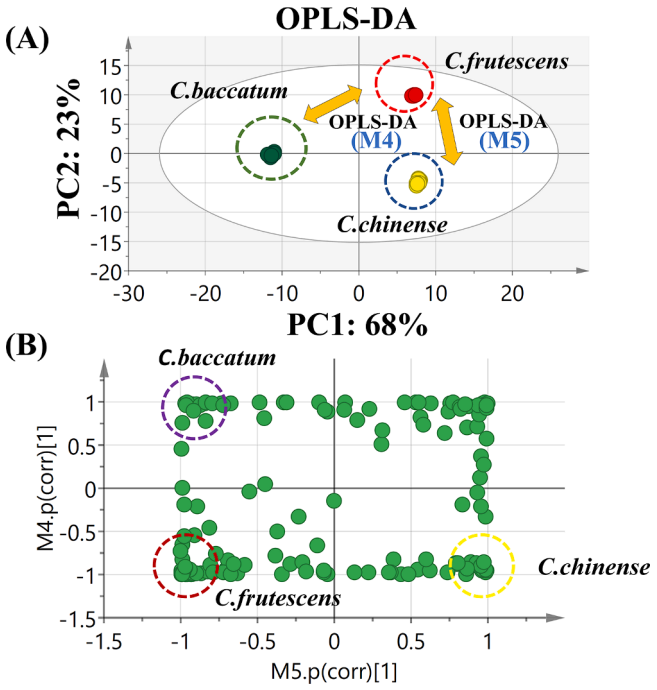

Fig. 4. The top (A) it is OPLS-DA of class species for samples of chili peppers (3 species) of data set of Chincha location by normalization EigenMS, (R^2 = 0.976), in B, 1) SUS-plot class species, (OPLS-DA, M4 = *C. baccatum* vs *C. frutescens*), (OPLS-DA, M5 = *C. frutescens* vs *C. chinense*). The diagram has been obtained by comparing the OPLS-DA, M4 and M5 models. In the vertices of the diagram, the biomarkers for each species were identified.

species of Chincha samples. The quality of the OPLS-DA model was evaluated with $R^2X = 0.998$ and $Q^2Y = 0.998$ values greater than 0.7 expressed a good quality of fit and prediction, respectively (Mickiewicz et al., 2015). The CV-ANOVA value confirmed the quality of this model (p-value = 3.14075e-0.35) (Eriksson, Trygg, and Wold, 2008). Then, OPLS-DA analyses were carried out facing two categories (2 species, Fig. 4A), to construct SUS-plot analyzes for identification of biomarkers (Fig. 4B). Fig. 5A shows the hierarchical analysis diagrams of Cluster via Heatmap of the relative areas of the 15 biomarker metabolites for each species (Table 2) and importance values (VIP) from biomarkers in Fig. 5B is showed, VIP < 1 is considered as the potential specific biomarkers.

Similarly, the samples were analyzed by accessions. The PCA of the accessions class by MS-CleanR approach is seen in Fig. S7. Figure S8 and S9 show the results of normalization by EigenMS of the original data with dimensions [182x66] columns × rows using the NOREVA platform, allowing to eliminate 17 "errors" of unknown origin, obtaining a final matrix of dimensions [165x66]. The normalized table was exported to the SIMCA-P software to perform the multivariate OPLS-DA analysis using Pareto scaling as shown in Fig. S9 (supplementary information).

From the OPLS-DA model by normalized accessions, two OPLS-DA models were constructed to determine the biomarkers that discriminate the accessions of the same species from the Chincha Area that is observed in Fig. S12: ayucllo vs tomatito rojo and miscucho vs dulce rojo. S-plot analysis was performed for OPLS-DA model respectively and discriminant biomarkers were identified as seen in Table S5 (Supplementary information).

From table S5, 20 biomarkers stand out for the discrimination of the accessions of which nine biomarker metabolites have an unknown structure: three unknowns from tomato rojo, four unknowns from

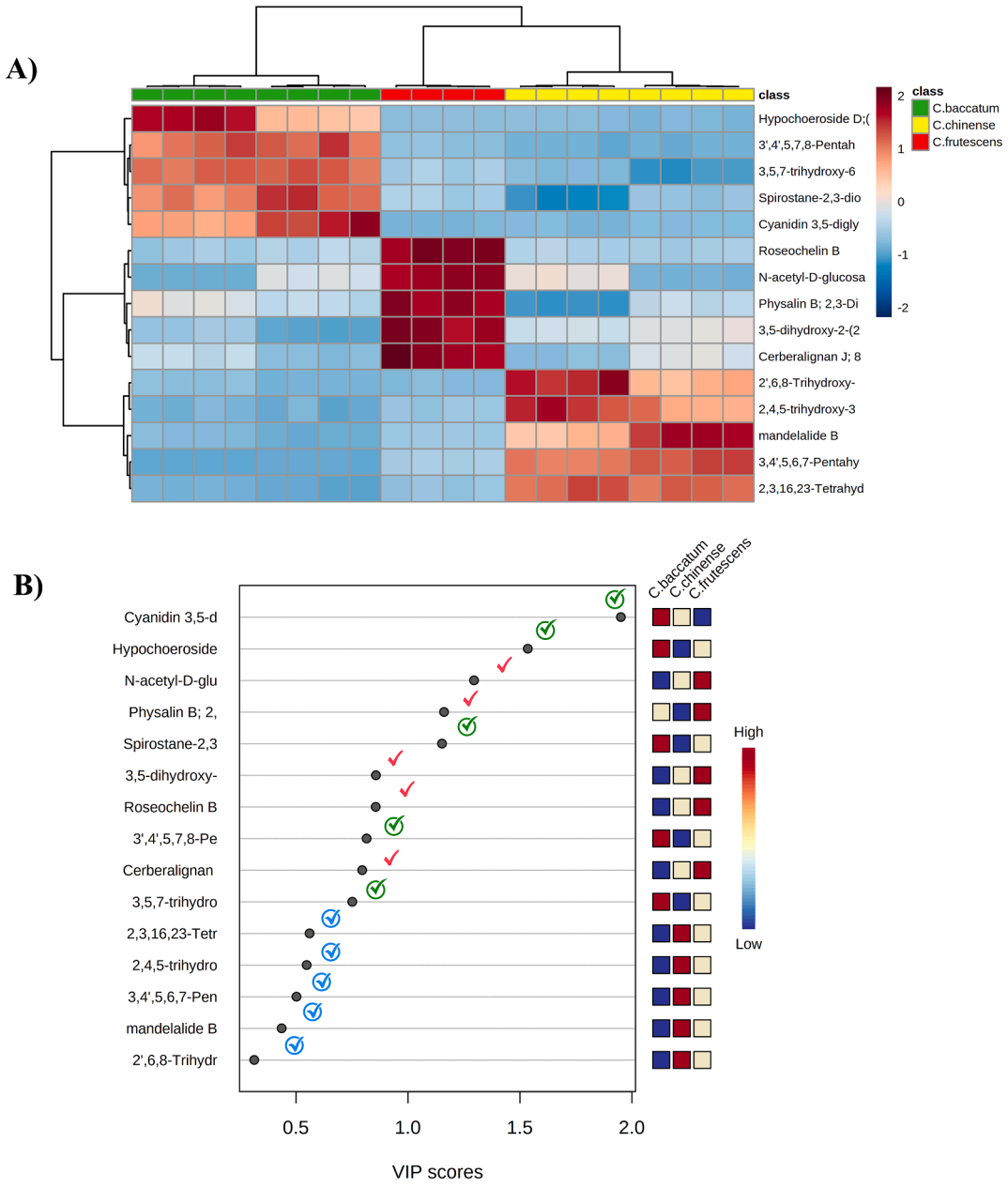

Fig. 5. In A) Cluster hierarchical analysis via Heatmap of biomarker metabolites from the PLS-DA model of the species class of the whole chili fruit from the Chincha location, in B) importance values (VIP) from biomarkers. The tick marks in green (*C. baccatum*) and red (*C. frutescens*) are the most contributing variables in the discrimination of the species in the PLS-DA model, VIP greater than 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ayucllo and two unknown miscucho respectively. This indicates the great scientific work that remains to be done to elucidate these chemical structures.

To validate the capacity of the model to classify species independently from collection area, samples from Huaral and Chiclayo were used as test set. The mean predicted class probabilities of each accession sample was measured for their respective species. A value above 0.5 indicates a high level of confidence in the predicted class (Fig. 6). The data was taken from the SIMCA-P classification list (supplementary information, Tables S1 and S2). The results of the bar chart show the predicted class probabilities of the OPLS-DA classification with values greater than 0.5.

The same procedure was used to validate the biomarkers of the accessions. Samples from Huaral and Chiclayo were used as a test set to validate the model's ability to classify accessions. A value greater than 0.5 indicates a high level of confidence in the predicted class (Fig. S10, Supplementary Information). Data were taken from the SIMCA-P classification list (Supplementary information, Tables S3 and S4).

4. Discussion

The objective of this study was to use metabolomic as a complementary tool to molecular tools and classical taxonomy visual descriptors for the discrimination of *C. baccatum*, *C. chinense* and *C. frutescens* species by identifying their biomarkers. To do this, the chili peppers from the town of Chincha were chosen to annotate biomarkers of each species. To verify the ubiquity of the biomarker metabolites in the three species studied, it was necessary to validate the metabolic fingerprint model of the five accessions of chili peppers from Chincha in chili fruits obtained from plants grown in the localities of Chiclayo and Huaral. In the same way, the possibility of discriminating accessions of the same species was explored.

INIA has an important collection of chili peppers and rocotos, collected in different regions of Peru, where morphological, genetic and physicochemical studies have been carried out, but not metabolomics (Ministerio del Ambiente 2015; INIA, 2018). Thus, this metabolomic study will contribute to the valorization of germplasm of Peruvian Capsicum. Chili peppers can be characterized and valued from the metabolic profile of the accessions studied through the composition of metabolites for their inclusion in genetic improvement programs and development of new cultivars.

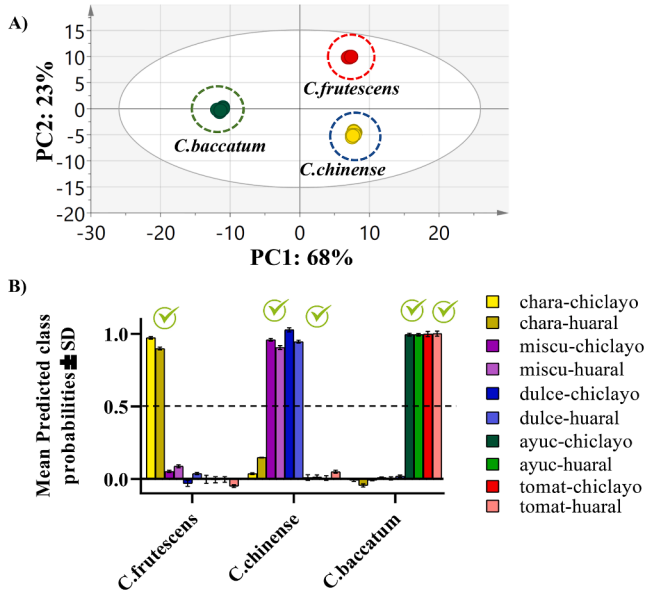

Fig. 6. OPLS-DA model classification for species as Y. A) OPLS-DA score plot from Chincha accessions with species as Y. B) The bar plot indicates the mean predicted class probabilities of each accession sample in their respective species. A value above 0.5 indicates a high level of confidence in the predicted class. Bar diagrams were made using GraphPad Prism 8 software.

The change of center of origin in itself already subjected the three species to a process of adaptation. This leads to the genotype and phenotype expression changing depending on environmental conditions. The expression of the phenotype could also be affected by the agricultural management of the chili peppers, since the chili peppers from Chincha had an organic agricultural management and the chili peppers from Chiclayo and Huaral, a conventional agricultural management. The sowing of the peppers was carried out in seedlings. Pests and diseases are the most significant problems affecting the crop. These include nematodes, fungi and pests that limit production and therefore require cultural practices that begin with the preparation of the crop, fertilization or compost. These conditions are difficult to manage and produce great variability in analytical results that adversely influence the metabolomic study. However, these difficulties were overcome, with the consequent identification of 14 biomarkers that discriminated the three species.

UHPLC-HRMS chromatography is an analytical approach that has been selected for this metabolomic study for its speed of analysis and precision for the determination of molecular masses, where there have been cases of use that have been successful in metabolomics studies of chili peppers. For example, the study carried out by Felipe Cervantes-Hernández (Cervantes-Hernández et al., 2019) established by means of UHPLC-ESI-HRMS metabolomic analysis the distribution of metabolites in the different parts of the fruit of the Tabasco chili pepper (*Capsicum frutescens L.*) when estimating the profiles of metabolites of the pericarp, placenta and seeds of the fruits. In another study by Sergio Barbosa et., al, the feasibility of non-targeted UHPLC-HRMS fingerprints as chemical descriptors was evaluated to address the classification and authentication of paprika samples. The obtained Untargeted UHPLC-HRMS fingerprints were subjected to unsupervised principal component analysis (PCA) and supervised partial least squares regression discriminant analysis (PLS-DA) to study the discrimination and classification of the sample. In this study it was concluded that the Untarget UHPLC-HRMS fingerprints Metabolomics were able to discriminate the different varieties of paprika (Barbosa et al., 2020).

Approximately 80–90% of the total capsaicinoids content of chili pepper is represented by capsaicin and Dihydrocapsaicin (Kulkarni et al., 2017). The most important differences between capsaicinoids are their relative pungency, the length of the branched chain, the point of branching and the amount of unsaturation in the side chain. Generally, capsaicin and Dihydrocapsaicin are almost 10 times hotter than other capsaicinoids (Kulkarni et al., 2017). The capsaicin / Dihydrocapsaicin ratio is 1: 1 or 2: 1 respectively (Giacalone, Forfori, and Giunta, 2015; Hayman and Kam, 2008). Capsaicin (*trans*-8-methyl-*N*-vanillyl-6-nonenamide) has a chemical structure consisting of a vanilloid group (an aromatic ring with a hydroxyl group and a methyl group), along with a long hydrocarbon chain and a polar amide group. The vanilloid group is common among other natural compounds in the so-called vanilloid family, such as vanillin, eugenol, and zingerone, and determines biological activity (Hayman and Kam, 2008).

Currently, chili pepper extract and its bioactive compounds are known to have a variety of pharmacological effects, such as antibacterial, antioxidant, analgesic, and anti-inflammatory effects. However, the effects of the main ingredient in chili peppers, capsaicin, is still controversial. It has potential biological activities at low concentrations, but tends to produce adverse effects at high concentrations (Hayman and Kam, 2008; Srinivasan, 2016). Around the world, there are about 2.5 billion people who eat spicy foods, and the spicy taste has become a global dietary trend. Capsaicin, the main hot component in chili peppers, has not been clearly studied for its potential effect on gastrointestinal health. Since capsaicin-induced gastrointestinal malaise is dose-dependent, future research should further analyze the different experimental models to clarify these correlations (Xiang et al., 2021). Plant

products of natural origin, such as chili peppers, have activity against *Helicobacter pylori*. Gastritis is considered one of the greatest health challenges, with *H. pylori* responsible for more than eighty percent of active chronic gastritis. Antibacterial activity of capsaicin has been found in vitro along with synergistic effects with other drugs. When combining capsaicin with metronidazole, for example, the inhibitory concentration values showed a synergistic effect. Therefore, capsaicin possesses promising anti-*H. pylori* bioactivity and synergistic activity when combined with metronidazole, but more research is needed to examine the mechanisms by which the biological activity was positive. In addition, it is necessary to ensure its activity against *H. pylori* in vivo and in clinical settings (Tayseer et al., 2020).

Also, the species of the genus Capsicum from Peru offer a wide range of sensory attributes. This is explained by its great diversity due to its different centers of origin, shapes, sizes, colors and a wide range of characteristics such as aromas, flavors and pungency. Of these five accessions studied, odors of citrus, passion fruit, oregano, herb, fruity, Apple and flavors of tomato, sweet, bell pepper and acidity have been characterized by a trained sensory panel (Patel et al., 2016).

According to the chromatographic profile, sugars and organic acids have been found in the peppers. The concentration of sugars and organic acids in peppers affect the flavor of the fruit. In addition, carbohydrates promote the stability of ascorbic acid and therefore improve the content of vitamins (Jarret et al., 2009).

The most abundant compounds identified in the five chili pepper accessions are flavonoids (Fig. S7, supplementary information). Flavonoids play a central role in various facets of plant life, especially in the interactions between the plant and the environment. These defend plants against various biotic and abiotic stresses, including ultraviolet radiation, pathogens, and insect pests (Treutter, 2006).

Among the flavonoids, there is an anthocyanin that was identified as a biomarker. Anthocyanins are very widespread pigments in nature that are responsible for the colors of fruits. Being native or wild chili peppers, they could develop defense mechanisms. In addition, these compounds can act as repellants for herbivores and parasites (E. T. Johnson, Berhow, and Dowd, 2008), attracting pollinators and seed dispersers, and protecting plants against biotic and abiotic stress (Naing and Kim, 2021).

In a study of agricultural production of carrots carried out by Cubero (Cubero-Leon, De Rudder, and Maquet, 2018), an OPLS-DA discriminant model was developed to predict the type of agricultural management from a data of four consecutive years. They were able to discriminate the type of organic or conventional agricultural production. Potential candidates for discriminant biomarkers were selected using the S-plot generated by the OPLS-DA models. The main difference in the present work of native peppers and that of carrots lies in the years of production. They have worked with crops of only one year of production. We do not work with peppers from other seasons due to economic issues since the production of peppers for this study comes from a pilot study of accessions and not varieties of peppers already domesticated. Our five accessions are still under agronomic studies for their domestication and massification. Also, the dataset used in the present study did not allow the discrimination based on agricultural practice (one crop of organic production and two crops of conventional production were used) since at least three repetitions are needed to build the model (three consecutive years of pepper harvests). It is also not possible to have a prediction model for the harvest area (Chincha, Chiclayo and Huaral), since information is only available for only one year of production. However, to discriminate species and accessions it was possible since there are at least three repetitions of samples of native peppers (three harvest areas).

The three species, being of the same Capsicum genus, express the same secondary metabolites commonly found in chili peppers, for example, the capsaicinoids *cis*-Capsaicin and Dihydrocapsaicin were present in the five accessions, but they differed in their relative concentrations. As the genus Capsicum has a great biodiversity, with the existence of hundreds of different cultivars for each species, the complete study on the biosynthesis of secondary metabolites and their functions within the plant and its relationship with its environment is a very laborious and difficult to confirm. Despite this difficulty, it is known that environmental conditions have a great influence on the genotype profile, and the state of maturity of the fruit.

The metabolites found in the three species of chili peppers in the present study mostly include the main chemical classes such as flavonoids, carboxylic acids, steroids, phenols and Carbohydrates. From the metabolite profile, it has been observed that there is no pattern in their content and they present a wide variability of functions corresponding to the genotype.

A fundamental part of the metabolomic analysis was the annotation of the metabolites. The use of MS-CleanR in tandem with MSDIAL-MSFINDER is allowed filtering the data of pseudo metabolites and target signals of interest, which saved analysis time. Within the total of biomarker metabolites by species, the most abundant chemical classes are Flavonoids (five metabolites), carboxylic acid (two metabolites) and steroids (two metabolites). In the PCA (Fig. 3A), a strong separation between the three species is observed, an intra-class separation is also observed between the samples. The intra-class separations could be the result of phenotypic variation as a consequence of the interaction of the genome with agroecological conditions. This variation could be expressed in different levels of concentration of the metabolites, absence of metabolites and presence of other metabolites according to the locality in which the chili pepper was grown.

For the validation of the metabolic fingerprint, an external validation of the OPLS-DA model of the chromatographic data of chili peppers from Chincha was carried out. The validation of the OPLS-DA model was carried out with prediction sets composed of replicas of the experiment in different locations (Chiclayo and Huaral). From a practical point of view, the validity of a model needs to be verified to predict a given phenomenon from a new experimental data, in such a way that the certainty of the results is reliable. There are some examples that can put into context the meaning of what can be achieved through a validation. For example, in the quality control of Ginseng materials in the search for their authenticity, applying vibrational spectroscopy (NIR and MIR) as analytical techniques (Sandasi et al., 2016). Another application is reported in the study carried out by Marti (Marti et al., 2019), which focused on the prediction of the type of solvent used in the production of extracts of *Serenoa repens* of four commercial brands, prepared with three extraction solvents: hexane, ethanol and supercritical CO_2, which were predicted using an OPLS-DA model built with the chromatographic data. It was possible to discriminate the types of solvents used, highlight the biomarker metabolites and predict the type of solvent used by the commercial brands.

For this purpose, the results matrix of the Chincha locality was used as a training set to build the OPLS-DA prediction model. The results matrix of the Chiclayo and Huaral localities were used as external or test sets. The results of predicted class probabilities showed a probability value greater than 0.5, which indicates a high predictive capacity of the Chincha OPLS-DA model, which implies a high probability of incidence of biomarkers in any of the three growing areas.

5. Conclusions

The untargeted metabolomics based on UHPLC-HRMS allowed the identification of biomarker metabolites for the discrimination of the three species (*C. baccatum*, *C. chinense* and *C. frutescens*) and four accessions (tomatito rojo vs ayucllo and miscucho vs dulce rojo) of the INIA collection of native chili peppers. These biomarker compounds were mainly Flavonoids, Steroids, Terpenoids, Anthocyanin and Carboxilic acids. This study is exploratory and can serve as an alternative tool to molecular biology tools to solve problems related to the taxonomy of the genus Capsicum. The untargeted metabolomics allowed us to study the chemistry of the phenotype since it was based on the profiling by UHPLC-HRMS of the secondary metabolites. The native peppers are

rich in flavonoids that give flavor to most of the fruits with characteristic colors and characteristic aromas well-valued by the increasingly demanding consumers of differentiated products. Therefore, the information generated will be useful for future undertakings in the area of quality improvement of chili peppers with certain phenotypic traits of metabolite profiles for the generation of improved species.

CRediT authorship contribution statement

Fabio Espichán: Conceptualization, Data acquisition, Data curation, Resources, Software, Writing – original draft, Writing – review & editing. **Rosario Rojas:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. **Fredy Quispe:** Conceptualization, Writing – review & editing. **Guillaume Cabanac:** Conceptualization, Resources, Software, Writing – review & editing. **Guillaume Marti:** Conceptualization, Data acquisition, Data curation, Resources, Software, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to the Victor Alzamora Castro Graduate School, to the Franco-Peruvian School of Life Sciences for the scholarship awarded to carry out the Doctoral Thesis in Life Sciences and to the Faculty of Pharmacy of the Paul Sabatier University of Toulouse-France, for their support and assistance in the UHPLC-HRMS analyzes. This study was also funded by FONDECYT-CONCYTEC under contract 134-2015-FONDECYT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.132704.

References

Alonso, A., Marsal, S., & Julià, A. (2015). Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Frontiers in Bioengineering and Biotechnology, 3*(March). https://doi.org/10.3389/fbioe.2015.00023

Barbosa, S., Saurina, J., Puignou, L., & Núñez, O. (2020). Classification and Authentication of Paprika by UHPLC-HRMS Fingerprinting and Multivariate Calibration Methods (PCA and PLS-DA). *Foods, 9*(4). https://doi.org/10.3390/foods9040486

Bosland, P. W., & Votava, E. J. (2012). Peppers: Vegetable and Spice Capsicums. *Crop Production Science in Horticulture Series 22* (2nd ed.). Cambridge, MA: CABI.

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., et al. (2018). Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics, 14*(6). https://doi.org/10.1007/s11306-018-1367-3

Calderón-Santiago, M., Fernández-Peralbo, M. A., Priego-Capote, F., María, D., & de Castro, L. (2016). MSCombine: A Tool for Merging Untargeted Metabolomic Data from High-Resolution Mass Spectrometry in the Positive and Negative Ionization Modes. *Metabolomics, 12*(3). https://doi.org/10.1007/s11306-016-0970-4

Cervantes-Hernández, F., Alcalá-González, P., Martínez, O., & Ordaz-Ortiz, J. J. (2019). Placenta, Pericarp, and Seeds of Tabasco Chili Pepper Fruits Show a Contrasting Diversity of Bioactive Metabolites. *Metabolites, 9*(10). https://doi.org/10.3390/metabo9100206

Chervin, Justine, Pierre Perio, Nathalie Martins-Froment, Chiobouaphong Pharkeovilay, Karine Reybier, Françoise Nepveu, Nicolas Fabre, Thierry Talou, Valérie Bonzon-Ponnet, and Guillaume Marti. 2017. "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* 13 (8). 10.1007/s11306-017-1227-6.

Cubero-Leon, E., De Rudder, O., & Maquet, A. (2018). Metabolomics for Organic Food Authentication: Results from a Long-Term Field Study in Carrots. *Food Chemistry, 239*(January), 760–770. https://doi.org/10.1016/j.foodchem.2017.06.161

Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for Significance Testing of PLS and OPLS® Models. *Journal of Chemometrics, 22*(11–12), 594–600. https://doi.org/10.1002/cem.1187

Eshbaugh, W. H. (1983). The Genus Capsicum (Solanaceae) in Africa. *Bothalia, 14*(3/4), 1983. https://doi.org/10.4102/abc.v14i3/4.1252

Fraisier-Vannier, Ophélie, Chervin, Justine, Cabanac, Guillaume, Puech-Pages, Virginie, Fournier, Sylvie, Durand, Virginie, Amiel, Aurélien, et al. (2020). MS-CleanR: A Feature-Filtering Approach to Improve Annotation Rate in Untargeted LC-MS Based Metabolomics. *Analytical Chemistry, 92*(14), 9971–9981. https://doi.org/10.1021/acs.analchem.0c01594

Giacalone, M., Forfori, F., & Giunta, F. (2015). Chili Pepper Compounds in the Management of Neuropathic Pain. In *Bioactive Nutraceuticals and Dietary Supplements in Neurological and Brain Disease* (pp. 187–195). Elsevier, 10.1016/B978-0-12-411462-3.00020-5.

Gika, H. G., Theodoridis, G. A., Wingate, J. E., & Wilson, I. D. (2007). Within-Day Reproducibility of an HPLC−MS-Based Method for Metabonomic Analysis: Application to Human Urine. *Journal of Proteome Research, 6*(8), 3291–3303. https://doi.org/10.1021/pr070183p

Gika, H., Virgiliou, C., Theodoridis, G., Plumb, R. S., & Wilson, I. D. (2019). Untargeted LC/MS-Based Metabolic Phenotyping (Metabonomics/Metabolomics): The State of the Art. *Journal of Chromatography B, 1117*(June), 136–147. https://doi.org/10.1016/j.jchromb.2019.04.009

Hayman, M., & Kam, P. C. A. (2008). Capsaicin: A Review of Its Pharmacology and Clinical Applications. *Current Anaesthesia & Critical Care, 19*(5–6), 338–343. https://doi.org/10.1016/j.cacc.2008.07.003

Hernández-Verdugo, S., Dávila, P., & Oyama, K. (2017). Síntesis Del Conocimiento Taxonómico. *Origen y Domesticación Del Género Capsicum.* 10.17129/botsci.1583.

INIA. 2018. "INIA y Universidad Agraria consolidarán Colección Nacional de Germoplasma de ajíes peruanos – Instituto Nacional de Innovación Agraria." 2018. http://www.inia.gob.pe/2018-nota-128/.

Jarret, R. L., Berke, T., Baldwin, E. A., & Antonious, G. F. (2009). Variability for Free Sugars and Organic Acids in Capsicum Chinense. *Chemistry & Biodiversity, 6*(2), 138–145. https://doi.org/10.1002/cbdv.200800046

Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E., & Edwards, L. M. (2014). "Metabolomics Data Normalization with EigenMS". Edited by Andrew R. Dalby. *PLoS ONE, 9*(12), Article e116221. https://doi.org/10.1371/journal.pone.0116221

Patel, K., Calderon, R., Asencios, E., Vilchez, D., Marcelo, M., & Rojas, R. (2016). Agro-Morphological Characteristics and Sensory Evaluation of Native Peruvian Chili Peppers. *Journal of Agricultural Science and Technology B, 6*(3). https://doi.org/10.17265/2161-6264/2016.03.005

Kulkarni, Y. A., Suryavanshi, S. V., Auti, S. T., & Gaikwad, A. B. (2017). Capsicum: A Natural Pain Modulator. In *Nutritional Modulators of Pain in the Aging Population* (pp. 107–119). Elsevier. https://doi.org/10.1016/B978-0-12-805186-3.00009-6.

Li, B.o., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., et al. (2017). NOREVA: Normalization and Evaluation of MS-Based Metabolomics Data. *Nucleic Acids Research, 45*(W1), W162–W170. https://doi.org/10.1093/nar/gkx449

Libreros, D., van Zonneveld, M., Petz, M., Meckelmann, S. W., Rios, L., Amaya, K., et al. (2013). *Catálogo de Ajíes (Capsicum Spp.) Peruanos Promisorios Conservados En El Banco de Semillas Del INIA - Perú. Bioversity International.*

Luan, H., Fenfen Ji, Y.u., & Chen, and Zongwei Cai.. (2018). StatTarget: A Streamlined Tool for Signal Drift Correction and Interpretations of Quantitative Mass Spectrometry-Based Omics Data. *Analytica Chimica Acta, 1036*(December), 66–72. https://doi.org/10.1016/j.aca.2018.08.002

Marti, G., Joulia, P., Amiel, A., Fabre, B., David, B., Fabre, N., et al. (2019). Comparison of the Phytochemical Composition of Serenoa Repens Extracts by a Multiplexed Metabolomic Approach. *Molecules, 24*(12), 2208. https://doi.org/10.3390/molecules24122208

Meckelmann, S. W., Riegel, D. W., van Zonneveld, M. J., Ríos, L., Peña, K., Ugas, R., et al. (2013). Compositional Characterization of Native Peruvian Chili Peppers (Capsicum Spp.). *Journal of Agricultural and Food Chemistry, 61*(10), 2530–2537. https://doi.org/10.1021/jf304986q

Mickiewicz, B., Tam, P., Jenne, C. N., Leger, C., Wong, J., Winston, B. W., et al. (2015). Integration of Metabolic and Inflammatory Mediator Profiles as a Potential Prognostic Approach for Septic Shock in the Intensive Care Unit. *Critical Care, 19*(1), 11. https://doi.org/10.1186/s13054-014-0729-0

Ministerio del Ambiente, Perú. 2015. "CIISB | Centro de Intercambio de Informacion Sobre Seguridad de La Biotecnologia Del Peru | Catálogo de Pasaporte Del Germoplasma de Capsicum." 2015. http://bioseguridad.minam.gob.pe/publicaciones_notas/catalogo-pasaporte-coleccion-germoplasma-capsicum/.

Moscone, E. A., Scaldaferro, M., Grabiele, M., Cecchini, N., García, Y. Sánchez., Jarret, R., Daviña, J., Ducasse, D., Barboza, G., & Ehrendorfer, F. (2014). In *The Evolution of Chili Peppers (Capsicum-Solanaceae): A Cytogenetic Perspective.* https://doi.org/10.17660/ActaHortic.2007.745.5

Naing, A. H., & Kim, C. K. (2021). Abiotic Stress-induced Anthocyanins in Plants: Their Role in Tolerance to Abiotic Stresses. *Physiologia Plantarum, 172*(3), 1711–1723. https://doi.org/10.1111/ppl.13373

Orgogozo, V., Morizot, B., & Martin, A. (2015). The Differential View of Genotype-Phenotype Relationships. *Frontiers in Genetics, 6*. https://doi.org/10.3389/fgene.2015.00179

Ortiz, R., Delgado, F., de la Flor, G., & Alvarado, and José Crossa.. (2010). Classifying Vegetable Genetic Resources—A Case Study with Domesticated Capsicum Spp. *Scientia Horticulturae, 126*(2), 186–191. https://doi.org/10.1016/j.scienta.2010.07.007

Patel, K., Ruiz, C., Calderon, R., Marcelo, M., & Rojas, R. (2016). Characterisation of Volatile Profiles in 50 Native Peruvian Chili Pepper Using Solid Phase Microextraction-Gas Chromatography Mass Spectrometry (SPME–GCMS). *Food*

Research International, 89(November), 471–475. https://doi.org/10.1016/j.foodres.2016.08.023

Toquica, P., Sandra, F. R., Martínez, E., Duque, M. C., & Tohme, J. (2003). Molecular Characterization by AFLPs of Capsicum Germplasm from the Amazon Department in Colombia. *Genetic Resources and Crop Evolution, 50*(6), 639–647. https://doi.org/10.1023/A:1024429320771

Sandasi, M., Vermaak, I., Chen, W., & Viljoen, A. (2016). The Application of Vibrational Spectroscopy Techniques in the Qualitative Assessment of Material Traded as Ginseng. *Molecules, 21*(4), 472. https://doi.org/10.3390/molecules21040472

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed Minimum Reporting Standards for Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics, 3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

Tayseer, I., Aburjai, T., Abu-Qatouseh, L., AL-Karabieh, N., Ahmed, W., Al-Samydai, A., et al. (2020). In Vitro Anti-Helicobacter Pylori Activity of Capsaicin. *Journal of Pure and Applied Microbiology, 14*(1), 279–286. https://doi.org/10.22207/JPAM.14.1.29

Treutter, D. (2006). Significance of Flavonoids in Plant Resistance: A Review. *Environmental Chemistry Letters, 4*(3), 147–157. https://doi.org/10.1007/s10311-006-0068-8

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nature Methods, 12*(6), 523–526. https://doi.org/10.1038/nmeth.3393

Wahyuni, Y., Ballester, A.-R., Tikunov, Y., de Vos, R. C. H., Pelgrom, K. T. B., Maharijaya, A., et al. (2013). Metabolomics and Molecular Marker Analysis to Explore Pepper (Capsicum Sp.) Biodiversity. *Metabolomics, 9*(1), 130–144. https://doi.org/10.1007/s11306-012-0432-6

Walsh, B. M., & Hoot, S. B. (2001). Phylogenetic Relationships of Capsicum (Solanaceae) Using DNA Sequences from Two Noncoding Regions: The Chloroplast AtpB-rbcL Spacer Region and Nuclear Waxy Introns. *International Journal of Plant Sciences, 162*(6), 1409–1418. https://doi.org/10.1086/323273

Wildman, R. E. C. (Ed.). (2007). *Handbook of Nutraceuticals and Functional Foods* (2nd ed.). Boca Raton: CRC/Taylor&Francis.

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0—Making Metabolomics More Meaningful. *Nucleic Acids Research, 43*(W1), W251–W257. https://doi.org/10.1093/nar/gkv380

Xiang, Q., Guo, W., Tang, X., Cui, S., Zhang, F., Liu, X., et al. (2021). Capsaicin—the Spicy Ingredient of Chili Peppers: A Review of the Gastrointestinal Effects and Mechanisms. *Trends in Food Science & Technology, 116*(October), 755–765. https://doi.org/10.1016/j.tifs.2021.08.034

Zhigila, D. A., AbdulRahaman, A. A., Kolawole, O. S., & Oladele, F. A. (2014). Fruit Morphology as Taxonomic Features in Five Varieties of Capsicum Annuum L. Solanaceae." Edited by Peter J. de Lange. *Journal of Botany*. https://doi.org/10.1155/2014/540868