Metabolomics profile responses to changing environments in a common bean (*Phaseolus vulgaris* L.) germplasm collection

Running title: Common beans metabolomics across environments

Elsa Mecha[1,2], Guillaume L. Erny[3], Ana C. L. Guerreiro[4,5], Rodrigo P. Feliciano[6], Inês Barbosa[1], Andreia Bento da Silva[7], Susana T. Leitão[1], Maria Manuela Veloso[8], Diego Rubiales[9], Ana Rodriguez-Mateos[6,10], Maria Eduardo Figueira[7], Maria Carlota Vaz Patto[1], Maria Rosário Bronze[1,2,7*]

[1]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal; emecha@itqb.unl.pt (E.M.); inesalexandrabarbosa@gmail.com (I.B.); sleitao@itqb.unl.pt (S.T.L.); cpatto@itqb.unl.pt (M.C.V.P.); mbronze@itqb.unl.pt (M.R.B.)

[2]iBET, Instituto de Biologia Experimental e Tecnológica, Av. da República, Apartado 12, 2781-901 Oeiras, Portugal; emecha@ibet.pt (E.M.); mbronze@ibet.pt (M.R.B.)*

[3]LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200 – 465 Porto, Portugal; guillaume@fe.up.pt (G.L.E.)

[4]UniMS – Mass Spectrometry Unit, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal; aguerreiro@itqb.unl.pt (A.C.L.G.)

[5]UniMS – Mass Spectrometry Unit, iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal; aguerreiro@ibet.pt (A.C.L.G.)

[6]Division of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, University of Düsseldorf, D-40225 Düsseldorf, Germany; Feliciano@med.uni-duesseldorf.de (R.P.F.)

[7]Faculdade de Farmácia, Universidade de Lisboa, Av. das Forças Armadas, 1649-019 Lisboa, Portugal; abentosilva@ff.ulisboa.pt (A.B.S.); efigueira@ff.ulisboa.pt (M.E.F.); mrbronze@ff.ulisboa.pt (M.R.B.)

[8]INIAV, Instituto Nacional de Investigação Agrária e Veterinária, 2784-505 Oeiras, Portugal; mveloso.inrb@gmail.com (M.V.)

[9]IAS, Institute for Sustainable Agriculture, CSIC, Avda Menéndez Pidal s/n, 14004 Córdoba, Spain; diego.rubiales@ias.csic.es (D.R.)

[10]Department of Nutritional Sciences, School of Life Course Sciences, King's College London, SE1 9NH London, UK; ana.rodriguez-mateos@kcl.ac.uk (A.R.M.)

*Address correspondence to:

Maria Rosário Bronze, Ph.D.

Food Functionality and Bioactives Lab,

iBET, Instituto de Biologia Experimental e Tecnológica, Av. da República, Apartado 12, 2781-901 Oeiras, Portugal mbronze@ibet.pt

Abstract

Metabolomics is one of the most powerful -*omics* to assist plant breeding. Despite the recognized genetic diversity in Portuguese common bean germplasm, details on its metabolomics profiles are still missing. Aiming to promote their use and to understand the environment's effect in bean metabolomics profiles, 107 Portuguese common bean accessions, cropped under contrasting environments, were analysed using spectrophotometric, untargeted and targeted mass spectrometry approaches. Although genotype was the most relevant factor on bean metabolomics profile, a clear genotype x environment interaction was also detected. Multivariate analysis highlighted, on the heat-stress environment, the existence of higher levels of salicylic acid, and lower levels of triterpene saponins. Three clusters were defined within each environment. White accessions presented the lowest content and the coloured ones the highest levels of prenol lipids and flavonoids. Sources of interesting metabolomics profiles are now identified for bean breeding, focusing either on local or on broad adaptation.

Keywords: *Phaseolus vulgaris*, metabolomics, diversity, mass spectrometry, multivariate analysis, correlation-based network analysis

1. Introduction

In the plant kingdom, there is a vast diversity of metabolites, up to 1 000 000 compounds, characterized by distinct chemical structures and present in a large range of concentrations (Obata & Fernie, 2012).

These plant metabolites can be classified as primary and secondary metabolites. Although this classification has been considered ambiguous (since the primary metabolites can also participate in plant metabolism as secondary metabolites) traditionally the term primary corresponds to molecules involved in living organisms' growth and survival. The term secondary concerns metabolites formed from the primary ones exerting functions related to environmental conditions' adaptability, such as defense against biotic and abiotic stresses, signaling and metal transport.

These secondary metabolites, including phenolic compounds (described as the most representative secondary metabolites found in plants (Lin et al., 2016)), may also exert antioxidant, anti-inflammatory and anti-carcinogenic activities in animal and human health, possessing undeniable economical value for pharmaceutical, nutraceutical and agro-industries (Thirumurugan, Cholarajan, Raja, & Vijayakumar, 2018).

Plant metabolomics, as a systematic, untargeted profiling of plant metabolites involved in core essential functions and in plant interactions with their environment have been also used to access the natural variance in metabolite content between individual plants representing a powerful tool to assist the improvement of crops' compositional quality (Schauer & Fernie, 2006; Bueno & Lopes, 2020).

Metabolomics studies have been mostly applied to model plant species and major crops such as tomato, rice, maize, wheat (Wu et al., 2018; Shi et al., 2020). Only a few published articles reported the use of *–Omics* in the study of common bean's metabolomics profile (Hernández et al., 2009; Mensack et al., 2010; Perez de Souza et al., 2019).

Common bean (*Phaseolus vulgaris* L.) represent one of the major grain legumes consumed worldwide for its edible seeds and pods (Zhang, Yasmin, & Song, 2019), being an important source of dietary protein and metabolites with potential health promoting effects, e.g. phenolic compounds and terpenoids (Bueno & Lopes, 2020). Portugal as part of the Iberian Peninsula is considered a secondary

center of common bean genetic diversity (Santalla, Rodiño, & De Ron, 2002), with many bean landraces still in cultivation (Leitão, Dinis, Veloso, Šatović, & Vaz Patto, 2017).

The first study on common bean metabolomics, using a non-targeted metabolite profiling approach conducted by gas chromatography – mass spectrometry, characterized metabolite profile changes in common bean roots, under phosphorus deficient soil conditions (Hernández et al., 2009). A second study conducted only with six cultivars and not focusing on metabolites identification, associated small molecules to distinct common bean centers of domestication (COD) (Mensack et al., 2010). The most recent study dedicated to the metabolite profiling of different common bean organs (seedlings, roots, leaves, flowers, pods) established, through an integrative network analysis, the tissue and accession specific metabolic diversity (Perez de Souza et al., 2019).

Although the genetic diversity of Portuguese common bean germplasm has been extensively recognized (Leitão et al., 2017), so far no study has focused on this germplasm metabolite diversity and/or on the impact of the environment in common bean metabolomics profile diversity.

The present study aimed to overcome the lack of knowledge regarding the natural variance in metabolites content of Portuguese common beans, in particular, and on common beans' metabolome variability under challenging environmental conditions, in general. To fulfill these goals common bean dry seeds from a Portuguese germplasm collection (n=107 accessions), cropped under two contrasting environments (traditional *versus* heat stress), were studied.

Disclosing the common bean seeds metabolomics profile under contrasting environments will provide useful information to breeders focused on improving common bean crop yields and quality, as well as to farmers facing climate change. This information will be useful to understand the impact of the environment on the beans' metabolome and therefore to predict specific metabolite levels under different environmental conditions. This can have implications on some future cropping adopted measures (e.g. sun exposure, irrigation conditions) in order to obtain an adequate level of specific metabolites. Characterizing Portuguese common beans' metabolome will create the opportunity to introduce the Portuguese common beans into breeding programs with the aim of giving response to a multitude of challenges, such as future warming climate conditions, crop productivity, resilience to

biotic and abiotic stresses and the demand of processors and consumers for accessions with attractive nutritional, nutraceutical and sensorial characteristics.

2. Materials and Methods

2.1. Chemicals

Folin-Ciocalteu's phenol reagent, sodium carbonate (99%), *(+)*-catechin (98%), sodium nitrite (97%), aluminium chloride (99.9%), and vanillin (99%) were purchased from Sigma-Aldrich (St. Louis, USA). Sulphuric acid (95–97%) was purchased from Fluka (Seelze, Germany). Sodium hydroxide (98%) was purchased from Merck (Darmstadt, Germany). Methanol (99.9%) was purchased from Carlo Erba Reagents (Rodano, Italy). Acetonitrile for LC-MS Ultra Chromasolv was purchased from Honeywell Riedel-de Haën™ (Seelze, Germany). Milli-Q® water (18.2 MΩ.cm) was obtained in a Millipore – Direct Q3 UV System equipment (Molsheim, France). Formic acid (98%) was obtained from Carl Roth (Karlsruhe, Germany). Eluents A and B used for Q-Orbitrap were from Optima™ LC/MS Grade, Fisher Scientific (NH, USA). Gallic acid, protocatechuic acid, *p*-hydroxybenzoic acid, sinapic acid, catechin, epicatechin, caffeic acid, *p*-coumaric acid, *t*-ferulic acid, kaempferol, quercetin were obtained from Sigma-Aldrich Co. (Steinheim, Germany).

2.2. Plant material

A total of 107 different common beans (*Phaseolus vulgaris* L.) accessions was provided by the Research Unit of Biotechnology and Genetic Resources germplasm bank at INIAV (PRT05), Portugal, covering different morphological characteristics (seeds color and size) as well as different gene pool of origin, as described previously (Leitão et al., 2017), **Table S1**. The different accessions were cropped in two contrasting environments, a traditional common bean cultivation environment at Cabrela in Portugal (latitude – 38°52'6.816''N and longitude – 9°21'15.804''W) and a stressful environment at Córdoba in Spain (latitude – 37°53'29.58''N and longitude – 4°46'21.90''W), following a randomized complete block design with two blocks (two biological replicates per accession, each containing 20 plants). The two environments were characterized by different average

temperatures (18 – 21 °C, in Cabrela, and 15 – 32°C, in Córdoba), different average relative humidity (66 – 80%, in Cabrela, and 31 – 63%, in Córdoba) and different types of soil (eutric cambisol in Cabrela and fluvisol in Córdoba) (Mecha et al., 2021).

2.3. Samples' preparation and extraction

The mature dried seeds of the viable plants were collected and milled (Falling nº 3100 – Perten, Sweeden) to a particle size of 0.8 mm and stored at -20 °C, until further analysis. Extracts of the milled common bean seeds were prepared in duplicate (technical replicates), as previously described (Mecha et al., 2019). Briefly, one gram of dry whole seed flour was extracted with 20 mL of methanol: water (60:40, v/v) solution, followed by sonication for 60 minutes. After centrifugation at 420 x g during 15 minutes, the supernatant was collected and the final volume adjusted to 20 mL. The prepared extracts were filtered through a 0.22 µm 13 mm CA syringe filter (GE WhatmanTM, Malborough, MA, USA) and kept at -20 ºC, until analysis.

2.4. Total phenolic content and total flavonoids content

Total phenolic content (TPC) and total flavonoids content (TFC) were measured by spectrophotometric methods as previously described (Mecha, et al., 2019). For TPC, after testing for the appropriate dilution, 3.5 mL of diluted extract were mixed to 0.100 mL of *Folin-Ciocalteu's* reagent. Sodium carbonate solution (35% w/v), 0.400 mL, was added to the mixture 3 min after. The absorbance was measured, after keeping the mixture during one hour in the dark, at 725 nm, in a Spectrophotometer DU-70 (Beckman®, Brea, CA, USA). A blank of water was also prepared in the same conditions and the gallic acid used as the external standard in a concentration range of 1 – 6 mg/L. The final results were expressed as mg of Gallic acid equivalents (GAE) per g of flour dry weight.

For TFC, after testing the appropriate dilution, 1 mL of diluted extract was added to 4 mL of Milli-Q® water and 0.300 mL of sodium nitrite (5%, w/v). After 5 min, 0.300 mL of aluminum chloride (10%, w/v) was added and to complete the reaction, after 6 minutes, 2 mL of 1 M sodium hydroxide solution were added. Milli-Q® water was applied to complete a final volume of 10 mL. Absorbance was

measured in a Spectrophotometer DU-70 (Beckman®, Brea, CA, USA), at 510 nm, against water. *(+)*-Catechin was applied as the external standard in a concentration range of 20 to 100 mg/L and the final results expressed as mg of *(+)*-catechin equivalents (CE) per g of flour dry weight (DW). The moisture content (%) of the raw flour used in the present study was retrieved from (Mecha et al., 2021) and determined by Near Infrared (NIR) analyser (MPA; Bruker, Billerica, MA, USA).

2.5. Untargeted metabolomics analysis by Q Exactive™ Focus Hybrid Orbitrap

The analysis of metabolites by untargeted metabolomics, in common bean extracts, was achieved by Orbitrap high-resolution mass spectrometry using a Q Exactive™ Focus Hybrid Q-Orbitrap (Thermo Scientific, MA, USA). For metabolites separation a XBridge BEH C18 (130 Å, 3.5 µm, 2.1 x 150 mm) column (Waters, MA, USA) was used. The elution was ensured with a binary system consisting of 0.1% formic acid in water (eluent A) and 0.1% formic acid in acetonitrile (eluent B), at a constant flow rate of 400 µL/min, during 20 minutes. The following gradient elution programme was applied: gradual increase of eluent B percent (from 1% of B at 1 minute to 99% of B at 13 minutes), followed by a steady percent of 99% of B during 2 minutes. At 15 minutes, the percent of B eluent returned to the initial conditions (1% of B) in one minute (from 99% of B at 15 minutes to 1% of B at 16 minutes). These conditions were maintained during 4 minutes before the next analysis starting. The UHPLC system (Dionex UltiMate 3000, Thermo Scientific, MA, USA) was coupled with a Q-Orbitrap mass spectrometer equipped with an electrospray ionization source working in negative and positive modes. The data were acquired in full-MS scan mode (scan range from 75 – 1125 m/z) with a resolution of 70 000 (at 200 m/z), $1x10^6$ automatic gain control (AGC) and internal calibration with lock mass (112.98550 *m/z*). The identification of compounds was fulfilled by Data-dependent method (ddMS2). The 3 most intense ions were subjected to higher energy collisional dissociation, HCD, 17 500 resolution, 20, 40, 60 normalized collision energy (NCE) and $1x10^5$ AGC. The maximum injection time was set at 100 ms and 6 s of dynamic exclusion. A quality control (QC) sample was prepared as a pool of distinct common bean extracts obtained from different common bean accessions (n=32). The selected accessions were characterized by distinct colours, seed size and gene pool of

origin, representing 30% of the total number of analysed accessions. In the sequence analysis, the quality control (QC) was analysed at every 63 injections, maximum once per day.

2.5.1. Data processing, identification and relative quantification of compounds

The collected data were analysed using the Finnee2016 toolbox for untargeted metabolomics analysis (Erny, Acunha, Simó, Cifuentes, & Alves, 2016). For more details please see supplementary material for materials and methods, **MMS1**.

The final excel file included a total of 1122 compounds (defined by *m/z* values) aligned accordingly to the retention time for further statistical analysis. For each feature (compound) the coefficient of variation between the obtained areas of the QC samples was lower than 20%.

The final excel file was exported to MetaboAnalyst (version 4.0) freely available at https://www.metaboanalyst.ca/, for statistical analysis and metabolites selection (Chong et al., 2018). The data were log transformed and pareto-scaled. Multivariate analysis by partial least square-discriminant analysis (PLS-DA) allowed to select the most relevant compounds responsible by genotype, gene pool of origin and environment differences, based on values of variable importance in projection (VIPs) higher than 0.8, as reviewed elsewhere (Mehmood, Liland, Snipen, & Sæbø, 2012). A *Venn* diagram was performed by *Venny* 2.1 (freely available at https://bioinfogp.cnb.csic.es/tools/venny/ (Oliveros, 2007)) applied as a tool to quickly distinguish the compounds exclusively responsible by genotype, gene pool of origin or environment distinction and the ones shared by the different groups.

After confirming the mass of the most abundant isotopes using XCalibur software (Thermo Fisher Scientific, MA, USA), the compounds were identified using the Compound Discoverer software, version 2.1, (Thermo Scientific™, MA, USA).

Considering the complexity of *Phaseolus* spp. metabolism it is worthy to notice that a clear identification by untargeted metabolomics can be difficult to attain, since the availability of authentic standards is often limited and the online databases are frequently incomplete or inconsistent. Therefore, to increase accuracy, for the putative identification (annotation) of metabolites, multiple databases were used (Zhou, Wang, & Ressom, 2012). The proposed annotations were considered

acceptable if there was at least match with one of the online databases (m/zCloud and/or Chemspider); a mass accuracy, $\frac{Predicted\ m/z - Observed\ m/z}{Predicted\ m/z} x\ 1000\ 000, \leq 1$ ppm at least in one of the ionization modes; at least one fragment with relation signal-to-noise higher than three, different from the parent ion, in common with the described fragmentation pattern (m/zCloud match score indicating similarity between experimental and described fragmentation spectra and Fragment Ion Search (FISh) scoring algorithm explaining fragment ions based on literature defined chemical reactions) and/or the compound was previously identified in Plantae kingdom and preferentially related to Fabaceae family. Whenever possible, freely available databases and published articles were used for data comparison (**Table S2**). For reliable annotations and since m/zCloud is a curated database of high-resolution tandem mass spectra, primacy was given to the identification made by this database, namely to compounds with m/zCloud match score higher than 80.0 (**Table 1**). All the retained fragment ions were characterized by intensity values higher than 10 000 counts. Classification of compounds into SuperClass level, Class level and categories within the Class was automated using the web-based application ClassyFire as described elsewhere (Feunang et al., 2016), freely available at http://classyfire.wishartlab.com/ (Feunang et al., 2016). The relative quantification was conducted by comparison of the percent area of individual compounds considering the different analysed common bean accessions.

2.6. Targeted metabolomics by UPLC-Q-TOF-MS

For quantification of individual phenolic compounds, the common bean extracts were analysed by targeted metabolomics using UPLC-Q-TOF-MS, in an Agilent 6550 iFunnel Accurate-Mass Q-TOF MS (Agilent, Waldbronn, Germany) equipment, with commercially available standards, following the procedure described elsewhere (Feliciano, Boeres et al., 2016). For more details please see supplementary material for materials and methods, **MMS2**. The quantified compounds were identified by comparison with the retention time and *m/z* values of the standards. Contents were expressed as μg per gram of raw flour in DW. The quantified concentrations were higher than the MQL previously described (Feliciano, Mecha, Bronze, & Rodriguez-Mateos, 2016). As described above classification

of the targeted compounds was also conducted by application of a web-based application, ClassyFire, freely available at http://classyfire.wishartlab.com/ (Feunang et al., 2016).

2.7. Statistical analysis

Using IBM® SPSS® Statistics, version 22 (IBM®, NY, USA), normality assumption was tested for each analysed parameter (Kolmogorov-Smirnov test at a significance level of 1%, variable's distribution in histogram and normal Q-Q plots) and when necessary different transformations were tested (logarithmic, inverse, square root, cubic root, fourth root and fractional ratio transformations) in order to achieve residuals' normality. Levene's test was used to test homoscedasticity at a significance level of 1%. The impact of the accession, environment, as well as the accession x environment interaction factors, was tested by nested ANOVA at a significance level of 5%. Significant differences were defined by post-hoc Scheffe's test or Games-Howell test, depending, respectively, on the presence or absence of equal variances. Eta^2 (%) was used to analyse the contribution of the different factors for parameters' variability. The adjusted R^2 indicated the quality of the models. One-way ANOVA was applied to test significant differences in the studied parameters considering the environmental conditions as a fixed factor and, for each environment, to test the existence of significant differences among the morphological aspects of common bean seeds, such as seed coat colour and seed size, as well as the gene pool of origin, at a significant level of 1%.

Multivariate analysis by principal component analysis (PCA) was performed to explore accessions' spatial grouping. The number of retained components was based on the Kaiser's criterion, eigenvalues higher than one, and the retained components applied in articulation with cluster analysis (K-means cluster analysis) to predict clusters' membership. For multivariate analysis purposes only the analysed parameters with communalities above 0.4 in the retained components were considered. The number of clusters was defined by a percentage of explained variance higher than 50%. Leave-one-out cross validation was applied in SPSS to assess the accuracy of the classification model. The probability of membership was determined by discriminant analysis. To sharpen groups' separation and establish correlations between the studied parameters and the defined clusters Partial Least Square – Discriminant Analysis (PLS-DA) was applied using Unscrambler® X 10.4.1, Camo Analytics

Software (Oslo, Norway). To facilitate the visualization of differences between clusters a heat map was established considering the relative quantification of the annotated compounds by Orbitrap-MS and the quantified parameters (TPC, TOF and metabolites quantified by Q-TOF-MS). The data collected by untargeted and targeted metabolomics were processed using the Correlation Calculator for Metabolomics data, freely available in the Metscape website (Basu et al., 2017). Metabolites pairwise partial correlations were calculated using DSPC (Debiased Sparse Partial Correlation) in order to measure the association between two metabolites without the confounding effect of all other metabolites related to them (Basu et al., 2017).

3. Results and Discussion

Only few metabolomics studies have been dedicated to the qualitative and quantitative diversity in common bean dry seeds (Llorach et al., 2019) or to the environmental impact, e.g. site of growth, (Quiroz-Sodi, Mendoza-Díaz, Hernández-Sandoval, & Carrillo-Ángeles, 2018), in their metabolomics profiles. In order to enlarge the existent knowledge to increase the efficiency of common bean breeding and production, the present study was conducted with 107 different underexploited Portuguese common bean accessions cropped in two contrasting environments and metabolite profiles from the harvested common bean dry seeds were further analysed by spectrophotometric and LC-Mass spectrometry methodologies.

3.1. Metabolic diversity of common bean dry seeds

The majority of metabolites were identified in the negative ionization mode, since it allows improved sensitivity (ionization efficiency) and lower detection limits, as described by Liigand et al., 2017, for the majority of metabolites non-anthocyanins.

Although anthocyanins are the major metabolites responsible by common bean seeds colour, in black, red and speckled common beans (Choung et al., 2003; Kan et al., 2016), herein the anthocyanins were not explored, since the existent online libraries are still limited in anthocyanins annotations and the selected m/z values, by PLS-DA, as the most important compounds responsible by samples' discrimination were mostly non-anthocyanins metabolites. The annotation of metabolites was carried

out using available online databases by comparison with mass accuracy, MS spectra and MS/MS fragmentation spectra. By using Q-Orbitrap-MS, 70 compounds, **Table S2**, from a dataset of 827 selected compounds were annotated, **Figure S1**. For the compounds' selection, PLS-DA analysis of an initial dataset of 1122 compounds was performed considering the environment, the accession and the gene pool of origin as fixed factors. Only the compounds with VIP scores higher than 0.8 (Mehmood et al., 2012) were selected. As shown in *Venn's* diagram, only 35.6 % of the selected compounds were responsible for the common bean samples' discrimination considering accession, gene pool of origin or environmental conditions, **Figure S1.** From the 70 annotated compounds identified in the negative ionization mode only ten of them were concomitantly identified in the positive ionization mode, fifteen and four of the annotated compounds showed respectively in the negative and positive mode reliable annotations with a mass accuracy lower or equal than one ppm and a m/zCloud match score higher than 80.0, **Table 1**, **Figure S2**. As shown in **Table S2**, 42 compounds were tentatively described, for the first time, in Fabaceae species, namely in common bean. Although multiple databases were used for compounds identification (Zhou, Wang, & Ressom, 2012), compounds annotation was impaired by the quality of the MS spectra and MS/MS fragmentation spectra published online. In fact, the previous poor investment in the legume metabolomics research field (only 10699 articles of plant metabolomics from a universe of 1262205 articles dedicated to plants at the date of manuscript writing, 09 July 2021) (Pubmed-NCBI) has overall hampered compounds annotation. The annotated compounds, **Figure 1A**, **Figure S3**, were classified, accordingly to the web-based application, ClassyFire, into seven different superclasses: organoheterocyclic compounds; phenylpropanoids and polyketides; organic oxygen compounds; benzenoids; lipids and lipid-like molecules; nucleosides, nucleotides and analogues and into the superclass of organic acids and derivatives. Most of the newly described compounds belonged to phenylpropanoids and polyketides as well as to lipids and lipid-like molecules superclasses. The phenylpropanoid and polyketides superclass was the one with higher diversity of compounds. This vast superclass of compounds comprises the largest pool of secondary metabolites, representing 20% of the total carbon in biosphere (Yu & Jez, 2008). Characterized by an aromatic ring linked to a three-carbon propene chain these compounds derived from deamination of phenylalanine (Fraser &

Chapple, 2011), **Figure S4**. With strong effects on plant growth and development, these compounds are also involved in the plant response to biotic and abiotic stresses, contributing to plant environmental adaptability and survival (Vogt, 2010). As shown in **Figure 1B**, the phenylpropanoid and polyketides superclass shares with the benzenoids superclass several metabolic pathways including the alkaloids and terpenoids biosynthesis. Additionally, phenylpropanoid and polyketides participate through the AMPK signalling pathway, **Figure 1B**, on downregulation of processes such as gluconeogenesis, lipid and protein synthesis, promoting fatty acid oxidation and autophagy, which may have interest for the treatment of type II diabetes, obesity and cancer (Jiménez-Sánchez et al., 2017; Thirumurugan et al., 2018). Within this superclass, the flavonoids class was the most abundant with a total of 21 tentatively identified compounds. Flavonoids are known to play several key roles in plants, contributing for the establishment of symbiotic relationships between plants and microorganisms, as well as in plant survival through the action of compounds that may induce insects and herbivores repelling and/or pollinators' attractiveness, e.g. anthocyanins (Ghasemzadeh & Ghasemzadeh, 2011).

The second most abundant superclass was the one named as lipids and lipid-like molecules, which included a total of 21 compounds. Into this superclass, the prenol lipids class was one of the most diversified classes with a total of 15 tentatively annotated compounds. Eleven of the 15 compounds were classified as triterpene saponins and triterpenoid compounds. Triterpenes are ubiquitous compounds in the plant kingdom, comprising six isoprene units in their structure. They can act as signalling molecules or as in the case of glycosylated triterpenes (saponins) as protecting compounds against pathogens (Thimmappa, Geisler, Louveau, O'Maille, & Osbourn, 2014). Triterpenes can be biosynthesized through the cytosolic mevalonate (MVA) pathway or alternatively by the plastidial non-mevalonate pathway (2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate), MEP/DOXP pathway, **Figure S4**. Several factors such as light and heat stress affect metabolic routes involved in the delivery and/or competition for carbon precursors (Liu et al., 2019), including the lipid and phenylpropanoid biosynthesis, **Figure S4**. Although a possible connection between the metabolic routes of terpenoids and phenylpropanoids exist, further studies are required to understand the regulation of both pathways (Tholl, 2015).

The benzenoids superclass includes compounds described in Kegg and MetaCyc databases as metabolites involved in the shikimate pathway, which participate in the synthesis of compounds with several essential roles in plant physiology (e.g. hormones, folate, amino acids and secondary metabolites biosynthesis). Involved in siderophore group biosynthesis, salicylic acid attenuates plant iron deficiency, especially in calcareous soil, where the availability of Fe (II) is impaired (Bakker, Ran, & Mercado-Blanco, 2014). Salicylic acid as a metabolite produced by the family Fabaceae has also an ecological role in the phytoremediation of contaminated soils participating in the degradation pathways of several soil contaminants/ pollutants such as polycyclic aromatic hydrocarbons, dioxins, toluene, naphthalene and bisphenol, **Figure 1B** (Hall, Soole, & Bentham, 2011; Saibu, Adebusoye, & Oyetibo, 2020).

The high qualitative diversity of metabolites described above was identified in the common bean accessions regardless of the cropping environmental conditions.

3.2. Effect of contrasting environments and accessions in common bean dry seeds metabolomics diversity

In the present study although no significant differences were found, under contrasting environments, in the overall total phenolic content (TPC) and total flavonoids content (TFC) determined by spectrophotometric methodologies, the use of hyphenated high-resolution separation techniques with accurate tandem mass spectrometry in conjugation with bioinformatic tools allowed the identification of specific metabolites that could be synergistically involved in common beans heat tolerance, **Table S2**. Under heat stress environmental conditions (Córdoba) the relative percent area of individual compounds classified as a pteridine derivative (Cp1), flavonoids (Cp16, Cp43, Cp51 and Cp52), isoflavonoid (Cp44), coumarin (Cp56), stilbene (Cp35), macrolide (Cp66), organo oxygen compounds (Cp3, Cp13), benzenoids (Cp9, Cp42, Cp45), fatty acyls (Cp65, Cp69), prenol lipids (Cp23, Cp38, Cp47, Cp54, Cp67, Cp68, Cp70) and carboxylic acids (Cp53) was significantly higher than in the milder traditional cropping environment of Cabrela, **Table S3**. The quantified benzenoid

compound (Cp72) and cinnamic acid (Cp77) also presented higher concentration in Córdoba than in Cabrela field trial, **Table S4**.

Multivariate analysis summarized the common bean dry seeds quantitative metabolomics diversity under contrasting environments (traditional, Cabrela *versus* stressful, Córdoba), **Figure 2**. The principal component analysis (PCA) established in a bi-dimensional space explained 58.4% of the total variance, and showed along the second principal component axis a clear separation in the metabolomics composition of common bean samples cropped under contrasting environmental conditions. Metabolites such as the reliably annotated, triterpene saponin (Cp60), **Table 1**, and the compounds annotated as fatty acyl Cp30 and purine nucleoside Cp26, **Table S2**, contributed to differentiate samples cropped in Cabrela from the ones cropped in Córdoba. The samples collected from Córdoba were projected mostly to the bottom of the representation as a consequence of the contribution of metabolites such as the reliably annotated salycilic acid (Cp42), the organo oxygen compound Cp13, the pteridine derivative Cp1, and the prenol lipid Cp47. The slightly higher contribution of environmental conditions (42 – 43%) to some metabolites variability (e.g. salycilic acid, Cp42, and pteridine compound, CP1) compared to the contributions attributed to accession or to accession x environment interaction, **Figure 3** and **Table S5**, unveiled the importance of these metabolites for common beans' survival and environmental adaptation.

The superimposition of the gene pool of origin in the bi-dimensional space indicated a potential trend in the metabolites abundance of common bean accessions classified into different gene pools of origin and cropped into different environmental conditions. On both environments, the common bean accessions classified into the Mesoamerican gene pool of origin were concentrated mostly in the left side of the first principal component axis, being generally characterized, by lower levels of phenylpropanoids and polyketides, total phenolic content (TPC) and total flavonoids content (TFC) than the majority of accessions belonging to other gene pools of origin. Although until now no metabolomics study has been performed in common bean accessions to understand the role of specific metabolites in common beans' heat tolerance, the obtained results are aligned with previous studies performed in other plant species. Under heat stress conditions, the development of reproductive organs and the nodulation process in legumes are impaired which decreases, respectively, the

401 fertilization and the nitrogen fixation (Bhandari et al., 2017). Despite the scarcity of metabolomics
402 studies reporting the specialized effect of individual metabolites in legumes adaptation to challenging
403 environmental stressful conditions, the activation of phenylpropanoid biosynthetic pathway under
404 abiotic stress is well recognized. As a consequence of the phenylalanine ammonia lyase increased
405 activity and polyphenol oxidase decreased activity, phenolic compounds responsible by plant
406 protection against reactive oxygen species accumulate in plant cells, enabling stress tolerance and
407 adaptation to challenging environments (Sharma et al., 2019). Triterpene saponins have also a key
408 role on plant growth and nodulation. Nevertheless, further investigation regarding their impact on
409 plant heat-stress tolerance is still required (Moses, Papadopoulou, & Osbourn, 2014). Regardless of
410 the high relative percent area of lipids and lipid-like molecules superclass, particularly prenol lipids,
411 **Table S3**, in Cabrela field trial, under heat-stress circumstances (Córdoba), the high percent area of
412 specific lipids and lipid-like metabolites (e.g. Cp23, Cp38, Cp47, Cp54, Cp67, Cp68, Cp70, Cp53)
413 anticipate their important contribute for the establishment and progress of nodulation counteracting
414 the adverse abiotic stress promoted by the temperature rising.
415 The role of metabolites such as pteridines derivatives (Cp1) has been described, in drought stress
416 conditions, as a co-factor for reactive oxygen species (ROS) scavenging enzymes, e.g. glutathione
417 reductase and NADPH-thiol reductase (Deng, Jin, Yang, Lin, & Zhang, 2014) and salicylic acid
418 (Cp42) showed, in wheat, the ability of improving photosynthesis under heat stress conditions through
419 enhancement of proline accumulation and inhibition of ethylene production (Khan, Iqbal, Masood,
420 Per, & Khan, 2013). Metabolomics studies conducted in other plant species such as in the carrots
421 showed the relevance of coumaric and caffeic acid as heat stress protectors (Commisso et al., 2016).
422 Notwithstanding the significantly high accession impact and the A x E interaction contribution
423 relatively reduced, < 20%, for the majority of the studied metabolites, **Figure 3** and **Table S5**, in just
424 a few metabolites such as the annotated compounds, azelaic acid (Cp33), hesperetin (Cp52),
425 succinyldisalicylic acid (Cp53), 6,7-Dihydroxy-4-methylcoumarin or isomer (Cp56) and ursolic acid
426 (Cp67), the contribution of A x E interaction to compounds' variability was ≥ 20%. These last
427 metabolites could be explored in breeding programs guided for the improvement of common bean

varieties' metabolomics profile only under particular homogeneous environments (varieties with local adaptation).

Cluster analysis of the PCA factor scores allowed the classification of common bean samples into six different groups, **Table S6**, which showed 98.3% of cross-validated grouped samples correctly classified.

The six groups explained 84.8% of the total variability and the difference between them sharpen by PLS-DA, **Figure 4**. The total explained variance of the different analysed parameters (predictors) and clusters (response) attributed to the two first components was respectively, 58% and 35%.

Although the common bean samples collected from Cabrela and Córdoba were clearly separated along the second component, the parameters with higher contribution for the first component were particularly relevant to distinguish common bean accessions within each environment. In fact, the spatial distribution of common bean samples from the same accession along the first principal component was quite similar within Cabrela and Córdoba field trials. On both environments, there were three clusters of common bean accessions (clusters 1, 3 and 4 for most of the Cabrela accessions and clusters 2, 5 and 6 for most of the Córdoba accessions) characterized by different metabolites abundance, **Figure 5**, **Table S7**. Clusters 2 and 3, located in the left side of the first component axis included, respectively, the common bean samples from Córdoba and Cabrela characterized by the lowest levels of TPC, TFC, catechin, epicatechin, annotated phenylpropanoids and polyketides (e.g. procyanidin C1, Cp7, quercetin-3-glucoside, Cp18, phloridzin or isomer, Cp16 and Cp20, quercetin, Cp46, plantagoside, Cp14, Calceolarioside B, Cp12 and phloretin, Cp49) and some specific lipids and lipid-like molecules (e.g. Cp61 and Cp69). The accessions included in these clusters were morphologically characterized by white seed coats. The observed differences in the TPC and TFC values among the diversity of Portuguese common bean seed coat colours, with white accessions showing the lowest content, **Table S8**, has been consistently documented previously in Portuguese (Mecha et al., 2019) accessions. The remaining clusters, morphologically characterized by coloured accessions, showed a huge qualitative and quantitative diversity of metabolites. In opposition to clusters 2 and 3, clusters 6 and 1, included respectively, the common bean accessions from Córdoba and Cabrela with the highest TPC, TFC and annotated phenylpropanoids and polyketides, **Figure 5**.

The richness of accessions, included in clusters 6 and 1, in metabolites mainly influenced by the accession factor (e.g. quercetin, Cp46, astragalin, Cp21, phloridzin or isomer, Cp20 and phloretin, Cp49) may have interest for future breeding programs focused in common beans broad adaptation. In fact, phenylpropanoids and polyketides (e.g. phenolic compounds) have been described as fundamental plant metabolites that improve the interaction between plant and rhizobacteria enhancing nutrient uptake and minerals mobilization. As antioxidants these metabolites can also protect plant cells from harmful conditions (e.g. UV radiation, temperature rising) that promote DNA mutations and ROS production (Sharma et al., 2019).

However there were huge differences in the abundance of some specific metabolites (e.g. Pteridine derivative, Cp1, benzenoids, Cp9 and Cp42, organooxygen compound, Cp13, and the prenol lipid, Cp23) which stood out in cluster 6, as an adaptation strategy to the environmental conditions. For future breeding programs focused in the development of varieties prone to heat-stress tolerance, breeders should focus mostly in accessions from cluster 6 particularly rich in metabolites, such as the benzenoid salycilic acid (Cp42), the organooxygen compound paeonoside (Cp13), the pteridine derivative Cp1 and the prenol lipid Cp47, with potential impact in common beans adaptability to specific environmental conditions. Cluster 1 included the accessions with the highest abundance of lipids and lipid-like molecules such as Cp61, Cp62, Cp63 as well as the highest abundance of benzenoids Cp73, Cp 28 and Cp37. While Clusters 1 and 6 were mostly composed by common bean accessions of large seed size (80% of the seeds), in the remaining clusters there were a predominance of medium and large seeds. Clusters 5 and 4 included, respectively in Córdoba and Cabrela, the common bean accessions with intermediate abundance of phenylpropanoids and polyketides. Cluster 4 stood out by the highest abundance of the triterpene saponins (Cp57, Cp59 and Cp60) and the triterpenoid Cp58, **Figure 5**. Regarding the gene pool of origin, the majority of common bean accessions with Mesoamerican origin were concentrated in clusters 2 and 5 as well as in clusters 3 and 4, which corresponded, overall, to clusters of common bean accessions characterized by the lowest or the intermediate abundance of the analysed parameters, as previously described.

3.3. Integrative approach to metabolite-metabolite interaction

As indicated by the partial correlations established between the analysed metabolites, **Table S9**, there were, strong interactions between the metabolites classified into the same superclass. The phenylpropanoids and polyketides superclass, which included the higher number of metabolites analysed, stood out by the high number of significant partial correlations, higher than 0.75, $p < 0.05$ (-log p >1.30). The compound annotated as kaempferol (Cp50) established partial correlations higher than 0.75 with astragalin (Cp21), 2'-acetylastragalin (Cp24), luteolin 7-O-(6-O-malonyl-β-glucoside) (Cp25), maesopsin (Cp34), and quercetin (Cp46). The last one was also highly and positively correlated to quercetin-3-β-D-glucoside (Cp18), taxifolin (Cp22) and protocatechuic acid (Cp72). The flavonol compounds, Cp50 and Cp46 share a common molecular backbone C6-C3-C6 consisting of two benzene rings (A and B) connected by a heterocyclic pyrane ring (C) and only few substitutions on the C ring (Cp21, Cp24, Cp18 and Cp22) or on the A (Cp25) rings explain the structure of the highly correlated metabolites. Cp34 and Cp72 classified, respectively, as an aurone flavonoid and as a hydroxybenzoic acid share with the flavonoids Cp50 and Cp46 the same biosynthetic pathway. As shown in **Figure S4**, aurone flavonoids and flavonols are synthesized via the phenylpropanoid pathway from the same precursor, p-coumaroyl-CoA. The dihydroxybenzoic acid, Cp72, can be produced via shikimate/chorismate or via phenylpropanoids (Widhalm & Dudareva, 2015), sharing with flavonoids, such as Cp46, a concomitant increase in their synthesis. This highly positive strong interaction was also observed between other annotated metabolites classified as benzenoids (aurantio-obtusin beta-D-glucoside, Cp37, homovanillic acid, Cp28 and syringic acid acetate, Cp9) and phenylpropanoids' metabolites (diosmin, Cp48, 5,7-dihydroxy-4-methylcoumarin or isomer, Cp40, sinapoyl D-glucoside, Cp27 and sinapic acid, Cp79), which supported the existence of common precursors in the metabolic routes responsible by the biosynthesis of metabolites classified into the two distinct superclasses.

The negative significant moderate partial correlations (-0.5 to -0.75, $p < 0.05$), between some metabolites classified into the phenylpropanoids and polyketides superclass (Cp50 *versus* Cp18; Cp50 *versus* kaempferol-3-O-rutinoside, Cp36) showed the complexity on the regulation of metabolites characterized by a similar backbone structure. Possible interconversions based on few substitutions at C, A and/or B rings are responsible by differences in the relative metabolites' proportion, in common

bean accessions. Such difference might be related to the natural variability in the flavanone-3-hydroxylase (F3H) enzymatic activity as well as in the flavonol UDP-glycosyltransferases among the different common bean accessions. As previously reported in safflower (*Carthamus tinctorius* L.) the existence of differential accumulation patterns of flavonoids could be attributed to different levels of F3H expression (Tu et al., 2016). F3H participates in flavonoid biosynthetic pathway acting in the 3-hydroxylation of flavanones into dihydroflavonols. Low expression of F3H could affect downstream the flavonol (e.g kaempferol) content (Tu et al., 2016). Moreover the qualitative diversity and the natural variability in the expression levels of flavonol UDP-glycosyltransferases (UGTs) could contribute to explain the accumulation of flavonols' glycosylated forms with concomitant reduction of flavonols upstream the flavonol biosynthetic pathway (Su et al., 2018). Besides the negative correlations between metabolites of the same superclass, there were also negative linear correlations established between metabolites of distinct superclasses, especially between compounds from benzenoids (e.g. Cp37) and lipids superclass (e.g. Cp38), as well as between phenylpropanoids (e.g. Cp66) and lipids superclass (e.g. Cp61), **Table S9**. A possible displacement of carbon precursors into the metabolic route of benzenoids and phenylpropanoid synthesis with a simultaneous downregulation in lipids and lipid-like molecules synthesis, **Figure S4**, (Vogt, 2010) could contribute for the observed differences in the proportion of metabolites belonging to distinct superclasses.

4. Conclusions

In the present study conducted with 107 Portuguese common bean accessions, cropped in two contrasting environments, 70 compounds, from an initial dataset of 1122 compounds, classified into seven different superclasses, were annotated. The compounds' annotation, performed by Q-Orbitrap-MS was impaired by the limited diversity of compounds described in available online libraries, as well as by the experimental and reported quality of MS spectra and MS/MS fragmentation spectra. Some of these compounds classified as phenylpropanoids and polyketides as well as lipids and lipid-like molecules were described for the first time in common bean extracts.

The multivariate data analysis showed the contribution of factors such as accession, environment and accession x environment interaction to metabolomics variability.

Despite the absence of significant differences in the total phenolic and total flavonoid contents determined in common bean accessions cropped under contrasting environments (traditional, Cabrela, Portugal *versus* heat stress, Córdoba, Spain), there were significant differences in individual metabolites content, namely in benzenoids (e.g. Cp42), lipids and lipid-like molecules (e.g. Cp57, Cp58, Cp59 and Cp60) and in organoheterocyclic compounds (e.g. Cp1).

Considering morphological traits such as seed coat colour, the coloured accessions highlighted, in both environments, as the ones with higher percent area of metabolites, including the ones classified as phenylpropanoids. Among coloured common bean accessions, two distinct clusters, based on metabolites abundance, were defined by PLS-DA analysis, within each environment. In relation to the gene pool of origin, accessions with Mesoamerican origin were mostly included in clusters characterized by lower percent area of metabolites than the accessions with an Andean or mixed origin. For the majority of the studied parameters, accession was the factor with the highest contribution ($Eta^2 > 50\%$) suggesting the high potential of the Portuguese common bean germplasm for future quality breeding programs. Common beans rich in metabolites mainly influenced by accession effect (e.g. astragalin, Cp21; quercetin, Cp46) will be interesting parental lines in breeding programs focused on the development of new varieties characterized by metabolomics profiles associated to higher potential nutraceutical effect, regardless of the environmental conditions where they will be cultivated (breeding for broad adaptation). Conversely, common bean accessions rich in metabolites with contents highly influenced by environmental conditions (e.g. salycilic acid, Cp42) may have interest for the breeding of varieties in challenging heat-stress environments (breeding for local adaptation). The metabolites pairwise partial correlations summarized the complex interactions established between the metabolites included into different superclasses, (defined in accordance to the ClassyFire web-based compounds classification) which contributed to elucidate shared metabolic pathways. Moreover, the list of detailed metabolites characterized in common bean accessions, and presented herein, may represent a starting point for future *in vitro* and *in vivo* studies focused on the

impact of single and multiple common beans' metabolites for human health, namely for the prevention of human non-communicable diseases.

Supplementary material:

MMS1. Untargeted metabolomics data processing (Finnee2016 toolbox); **MMS2.** Targeted metabolomics by UPLC-Q-TOF-MS; **List of References** (MMS1, MMS2, Table 1 and Table S2); **Table S1.** Morphological aspects of seeds and described gene pool of origin (Mesoamerican; Andean; Mixed) of Portuguese common bean accessions (Leitão, et al., 2017); **Table S2.** Tentative identification of metabolites in common bean accessions using *Compound Discoverer* software. The compounds tentatively identified, by m/zCloud and/or Chemspider (CS) identifications (ID), using the negative and/or positive electrospray ionization modes (ESI – and/or ESI+) are presented. The majority of tentative identifications were achieved in the negative ionization mode. For few compounds, the data collected in negative and positive ionization modes are presented. For more details about the references please consult the list provided in Supplementary material; **Table S3.** Comparison of compounds' relative quantification, %, (average ± standard deviation, SD) determined in common bean accessions, cropped under contrasting environmental conditions, using Orbitrap-MS. [a,b] significant differences ($p < 0.01$) *Below limit of quantification; **Table S4.** Total phenolic content (TPC) in mg GAE/g DW, total flavonoids content (TFC) in mg CE/g DW and quantification of individual metabolites (Cp71, Cp72, Cp73, Cp74, Cp75, Cp76, Cp77, Cp78, Cp79, Cp46 and Cp50), average ± SD, by UPLC-Q-TOF-MS in μg/g DW, determined in the Portuguese common bean accessions cropped in in contrasting environments (1, Cabrela and 2, Córdoba). [a,b] significant differences between average values *Below the limit of quantification (Feliciano, Mecha, et al., 2016); **Table S5.** Detailed information regarding the impact of accession (A), environment (E), block within environment (B(E)) and accession x environment (AxE) interaction in common beans' metabolites variability; **Table S6.** Classification of common bean accession in the different clusters (CL1 – CL6). Values in bold indicate probabilities of memberships higher than 0.5000; **Table S7.** Common bean clusters, per environment, considering the parameters relevant for the two first principal components (communalities higher than 0.4). All the results were expressed as average ± standard error of mean (SEM). For Cp72, Cp73, Cp74 and Cp76 the results were presented as μg/g DW. For total phenolic content (TPC) and total flavonoids content (TFC) as mg of gallic acid equivalents (GAE)/ g dry weight (DW) and mg of catechin equivalents (CE)/g DW, respectively. The remaining parameters were expressed as percent areas (%);**Table S8.** Total phenolic content (TPC) in mg GAE/g DW, total flavonoids content (TFC) in mg CE/g DW, relative quantification of metabolites' superclasses in % and quantification of phenolic compounds (Cp71, Cp72, Cp73, Cp74, Cp75, Cp76, Cp77,

Cp78, Cp79, Cp46 and Cp50) in μg/g DW, average ± standard deviation (SD), considering the white and coloured common bean accessions cropped in the two contrasting environments (Cabrela and Córdoba); **Table S9.** Partial correlations and correspondent significance (-log_p-value) determined between pairs of annotated metabolites by DSPC (Debiased Sparse Partial Correlation); **Figure S1.** Venn diagram showing the number of selected compounds related with differences in accessions, cropping environment and common beans gene pool of origin. The number of compounds shared by the different factors is shown in the intersection zones. The number underlined inside squares indicates the number of compounds with acceptable annotations; **Figure S2.** MS and MS^2 spectra of the annotated metabolites with mass accuracy (Δ mass) lower than 1 ppm and *mzCloud* match score higher than 80.0, in common bean accessions using *Compound Discoverer* software. MS^2 spectra of annotated compounds were compared to the described spectra in online libraries (m/z Cloud and ChemSpider); **Figure S3.** Molecular structure of identified common bean metabolites, organized into the different compounds' classes; **Figure S4.** Simplified representation of the metabolic pathways involved in common bean metabolites' synthesis (G-6-P, glucose-6-phosphate; Ribulose 5-P, ribulose 5-phosphate; Erythose-4-P, erithose-4-phosphate; Glyceraldehyde 3-P, glyceraldehyde 3-phosphate; Pentose-P, pentose-phosphate; MEP/DOXP, 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate ; VLCFA, very long chain fatty acids; L-Phe, L-Phenylalanine), adapted (Vogt, 2010).

5. References

Bakker, P. A. H. M., Ran, L., & Mercado-Blanco, J. (2014). Rhizobacterial salicylate production provokes headaches! *Plant and Soil, 382*(1), 1-16. https://doi.org/10.1007/s11104-014-2102-0.

Basu, S., Duren, W., Evans, C. R., Burant, C. F., Michailidis, G., & Karnovsky, A. (2017). Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. *Bioinformatics (Oxford, England), 33*(10), 1545-1553. https://doi.org/10.1093/bioinformatics/btx012.

Bhandari, K., Sharma, K. D., Hanumantha Rao, B., Siddique, K. H. M., Gaur, P., Agrawal, S. K., Nair, R. M., & Nayyar, H. (2017). Temperature sensitivity of food legumes: a physiological insight. *Acta Physiologiae Plantarum, 39*(3), 68. https://doi.org/10.1007/s11738-017-2361-5.

Bueno, P. C. P., & Lopes, N. P. (2020). Metabolomics to characterize adaptive and signaling responses in legume crops under abiotic stresses. *ACS Omega, 5*(4), 1752-1763. https://doi.org/10.1021/acsomega.9b03668.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S., & Xia, J. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research, 46*(W1), W486-W494. https://doi.org/10.1093/nar/gky310.

Choung, M-G., Choi, B-R., An, Y-N., Chu, Y-H., Cho, Y-S. (2003). Anthocyanin profile of Korean cultivated kidney bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, 51(24), 7040-7043. https://doi.org/10.1021/jf0304021.

Commisso, M., Toffali, K., Strazzer, P., Stocchero, M., Ceoldo, S., Baldan, B., Levi, M., & Guzzo, F. (2016). Impact of phenylpropanoid compounds on heat stress tolerance in carrot cell cultures. *Frontiers in plant science*, 7(1439). https://doi.org/10.3389/fpls.2016.01439.

Cytoscape consortium. Cytoscape. Network data integration, analysis and visualization in a box. Retrieved from https://cytoscape.org/. Accessed 13 September 2020.

de Mejía, E. G., Guzmán-Maldonado, S. H., Acosta-Gallegos, J. A., Reynoso-Camacho, R., Ramírez-Rodríguez, E., Pons-Hernández, J. L., González-Chavira, M. M., Castellanos, J. Z., & Kelly, J. D. (2003). Effect of cultivar and growing location on the trypsin inhibitors, tannins, and lectins of common beans

(*Phaseolus vulgaris* L.) grown in the semiarid highlands of Mexico. *Journal of Agricultural and Food Chemistry*, *51*(20), 5962-5966. https://doi.org/10.1021/jf030046m.

Deng, B., Jin, X., Yang, Y., Lin, Z., & Zhang, Y. (2014). The regulatory role of riboflavin in the drought tolerance of tobacco plants depends on ROS production. *Plant Growth Regulation, 72*(3), 269-277. https://doi.org/10.1007/s10725-013-9858-8.

Erny, G., Acunha, T., Simó, C., Cifuentes, A., & Alves, A. (2016). Finnee — A Matlab toolbox for separation techniques hyphenated high resolution mass spectrometry dataset. *Chemometrics and Intelligent Laboratory Systems, 155*, 138-144. https://doi.org/10.1016/j.chemolab.2016.04.013.

Feliciano, R. P., Boeres, A., Massacessi, L., Istas, G., Ventura, M. R., Nunes dos Santos, C., Heiss, C., Rodriguez-Mateos, A. (2016). Identification and quantification of novel cranberry-derived plasma and urinary (poly)phenols. *Archives of Biochemistry and Biophysics, 599*, 31-41. https://doi.org/10.1016/j.abb.2016.01.014.

Feliciano, R. P., Mecha, E., Bronze, M. R., & Rodriguez-Mateos, A. (2016). Development and validation of a high-throughput micro solid-phase extraction method coupled with ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry for rapid identification and quantification of phenolic metabolites in human plasma and urine. *Journal of Chromatography A, 1464*, 21-31. https://doi.org/10.1016/j.chroma.2016.08.027.

Feunang, D. Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., & Wishart, D. S. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics, 8*, 61.

Fraser, C. M., & Chapple, C. (2011). The phenylpropanoid pathway in *Arabidopsis*. *The arabidopsis book, 9*, e0152-e0152. https://doi.org/10.1199/tab.0152.

Ghasemzadeh, A., & Ghasemzadeh, N. (2011). Flavonoids and phenolic acids: Role and biochemical activity in plants and human. *Journal of Medicinal Plants Research*, 5. https://doi.org/10.5897/jmpr11.1404.

Hall, J., Soole, K., & Bentham, R. (2011). Hydrocarbon phytoremediation in the family Fabacea—A Review. *International Journal of Phytoremediation*, 13(4), 317-332. https://doi.org/10.1080/15226514.2010.495143.

Hernández, G., Valdés-López, O., Ramírez, M., Goffard, N., Weiller, G., Aparicio-Fabre, R., Fuentes, S. I., Erban, A., Kopka, J., Udvardi, M. K., & Vance, C. P. (2009). Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. *Plant physiology, 151*(3), 1221-1238. https://doi.org/10.1104/pp.109.143842.

Jiménez-Sánchez, C., Olivares-Vicente, M., Rodríguez-Pérez, C., Herranz-López, M., Lozano-Sánchez, J., Segura-Carretero, A., Fernández-Gutiérrez, A., Encinar, J. A., & Micol, V. (2017). AMPK modulatory activity of olive-tree leaves phenolic compounds: Bioassay-guided isolation on adipocyte model and in silico approach. *PloS one*, 12(3), e0173074-e0173074. https://doi.org/10.1371/journal.pone.0173074.

Kan, L., Nie, S., Hu, J., Liu, Z., Xie, M. (2016). Antioxidant activities and anthocyanins composition of seed coats from twenty-six kidney bean cultivars. *Journal of Functional Foods*,26, 622-631. https://doi.org/10.1016/j.jff.2016.08.030.

Khan, M. I. R., Iqbal, N., Masood, A., Per, T. S., & Khan, N. A. (2013). Salicylic acid alleviates adverse effects of heat stress on photosynthesis through changes in proline production and ethylene formation. *Plant Signaling & Behavior, 8*(11), e26374. https://doi.org/10.4161/psb.26374.

Kiralj, R., & Ferreira, M. M. C. (2009). Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society, 20*, 770-787.

Leitão, S. T., Dinis, M., Veloso, M. M., Šatović, Z., & Vaz Patto, M. C. (2017). Establishing the bases for introducing the unexplored Portuguese common bean germplasm into the breeding world. *Frontiers in Plant Science, 8*(1296). https://doi.org/10.3389/fpls.2017.01296.

Liigand, P., Kaupmees, K., Haav, K., Liigand, J., Leito, I., Girod, M., Antoine, R., & Kruve, A. (2017). Think negative: finding the best electrospray ionization/MS mode for your analyte. *Analytical Chemistry*; 89 (11), 5665-5668. https://doi.org/10.1021/acs.analchem.7b00096.

Lin, D., Xiao, M., Zhao, J., Li, Z., Xing, B., Li, X., Kong, M., Li, L., Zhang, Q., Liu, Y., Chen, H., Qin, W., Wu, H., & Chen, S. (2016). An overview of plant phenolic compounds and their importance in human nutrition and management of type 2 diabetes. *Molecules, 21*(10), 1374.

Liu, Y., Li, J., Zhu, Y., Jones, A., Rose, R. J., & Song, Y. (2019). Heat stress in legume seed setting: Effects, causes, and future prospects. *Frontiers in plant science, 10*(938). https://doi.org/10.3389/fpls.2019.00938.

Llorach, R., Favari, C., Alonso, D., Garcia-Aloy, M., Andres-Lacueva, C., & Urpi-Sarda, M. (2019). Comparative metabolite fingerprinting of legumes using LC-MS-based untargeted metabolomics. *Food Research International, 126*, 108666. https://doi.org/10.1016/j.foodres.2019.108666.

Mecha, E., Leitão, S. T., Carbas, B., Serra, A. T., Moreira, P. M., Veloso, M. M., Gomes, R., Figueira, M. E., Brites, C., Vaz Patto, M. C., & Bronze, M. R. (2019). Characterization of soaking process' impact in common beans phenolic composition: contribute from the unexplored Portuguese germplasm. *Foods* (Basel, Switzerland), 8(8), 296. https://doi.org/10.3390/foods8080296.

Mecha, E., Natalello, S., Carbas, B., da Silva, A. B., Leitão, S. T., Brites, C., Veloso, M. M., Rubiales, D., Costa, J., Cabral, M. dF., Figueira, M. E., Vaz Patto, M. C., & Bronze, M. R. (2021). Disclosing the nutritional quality diversity of Portuguese common beans—The missing link for their effective use in protein quality breeding programs. *Agronomy*, 11(2), 221. https://doi.org/10.3390/agronomy11020221.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62-69. https://doi.org/10.1016/j.chemolab.2012.07.010.

Mensack, M. M., Fitzgerald, V. K., Ryan, E. P., Lewis, M. R., Thompson, H. J., & Brick, M. A. (2010). Evaluation of diversity among common beans (*Phaseolus vulgaris* L.) from two centers of domestication using 'omics' technologies. *BMC Genomics, 11*(1), 686. https://doi.org/10.1186/1471-2164-11-686.

Moses, T., Papadopoulou, K. K., & Osbourn, A. (2014). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Critical reviews in biochemistry and molecular biology, 49*(6), 439-462. https://doi.org/10.3109/10409238.2014.953628.

Obata, T., & Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and molecular life sciences: CMLS, 69*(19), 3225-3243. https://doi.org/10.1007/s00018-012-1091-5.

Oliveros, J. C. (2007). Venny. An interactive tool for comparing lists with Venn's diagrams. Retrieved from https://bioinfogp.cnb.csic.es/tools/venny/. Accesed 25 May 2020.

Perez de Souza, L., Scossa, F., Proost, S., Bitocchi, E., Papa, R., Tohge, T., & Fernie, A. R. (2019). Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. *The Plant Journal, 97*(6), 1132-1153. https://doi.org/10.1111/tpj.14178.

Pubmed-NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed. Accessed 18 May 2020.

Quiroz-Sodi, M., Mendoza-Díaz, S., Hernández-Sandoval, L., & Carrillo-Ángeles, I. (2018). Characterization of the secondary metabolites in the seeds of nine native bean varieties (*Phaseolus vulgaris* and *P. coccineus*) from Querétaro, Mexico. *Botanical Sciences, 96*, 650-661.

Saibu, S., Adebusoye, S. A., & Oyetibo, G. O. (2020). Aerobic bacterial transformation and biodegradation of dioxins: a review. *Bioresources and Bioprocessing, 7*(1), 7. https://doi.org/10.1186/s40643-020-0294-0.

Santalla, M., Rodiño, A., & De Ron, A. (2002). Allozyme evidence supporting southwestern Europe as a secondary center of genetic diversity for the common bean. *Theoretical and Applied Genetics, 104*(6), 934-944. https://doi.org/10.1007/s00122-001-0844-6.

Schauer, N., & Fernie, A. R. (2006). Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science, 11*(10), 508-516. https://doi.org/10.1016/j.tplants.2006.08.007.

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., & Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules (Basel, Switzerland), 24*(13), 2452. https://doi.org/10.3390/molecules24132452.

Shi, T., Zhu, A., Jia, J., Hu, X., Chen, J., Liu, W., Ren, X., Sun, D., Fernie, A. R., & Chen, W. (2020). Metabolomics analysis and metabolite-agronomic trait associations using kernels of wheat (*Triticum aestivum*) recombinant inbred lines. *The Plant Journal, 103*(1), 279-292. https://doi.org/10.1111/tpj.14727.

Su, X., Wang, W., Xia, T., Gao, L., Shen, G., Pang, Y. (2018). Characterization of a heat responsive UDP: Flavonoid glucosyltransferase gene in tea plant (*Camellia sinensis*). *PLoS One*, 13(11), e0207212. https://doi.org/10.1371/journal.pone.0207212.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., & Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annual Review of Plant Biology, 65*(1), 225-257. https://doi.org/10.1146/annurev-arplant-050312-120229.

Thirumurugan, D., Cholarajan, A., Raja, S. S. S., & Vijayakumar, R. (2018). An introductory chapter: Secondary metabolites. In R. Vijayakumar & S. S. S. Raja (Eds.), *Secondary Metabolites - Sources and Applications*.

Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Advances in biochemical engineering/biotechnology, 148*. https://doi.org/10.1007/10_2014_295.

Tu, Y., Liu, F., Guo, D., Fan, L., Zhu, Z., Xue, Y., Gao, Y., Guo, M. (2016). Molecular characterization of flavanone 3-hydroxylase gene and flavonoid accumulation in two chemotyped safflower lines in

response to methyl jasmonate stimulation. *BMC Plant Biology*, 16(1), 132. https://doi.org/10.1186/s12870-016-0813-5.
Vogt, T. (2010). Phenylpropanoid biosynthesis. *Molecular Plant, 3*(1), 2-20. https://doi.org/10.1093/mp/ssp106.
Widhalm, J. R., & Dudareva, N. (2015). A familiar ring to it: biosynthesis of plant benzoic acids. *Molecular plant, 8*(1), 83-97. https://doi.org/10.1016/j.molp.2014.12.001.
Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., Méret, M., Keurentjes, J. B., Nikoloski, Z., Fernie, A. R., Willmitzer, L., & Brotman, Y. (2018). Mapping the *Arabidopsis* metabolic landscape by untargeted metabolomics at different environmental conditions. *Molecular Plant, 11*(1), 118-134. https://doi.org/10.1016/j.molp.2017.08.012.
Yu, O., & Jez, J. M. (2008). Nature's assembly line: biosynthesis of simple phenylpropanoids and polyketides. *The Plant Journal, 54*(4), 750-762. https://doi.org/10.1111/j.1365-313X.2008.03436.x.
Zhang, H., Yasmin, F., & Song, B.-H. (2019). Neglected treasures in the wild — legume wild relatives in food security and human health. *Current Opinion in Plant Biology, 49*, 17-26. https://doi.org/10.1016/j.pbi.2019.04.004.
Zhou, B., Wang, J., & Ressom, H. W. (2012). MetaboSearch: tool for mass-based metabolite identification using multiple databases. *PloS one, 7*(6), e40096-e40096. https://doi.org/10.1371/journal.pone.0040096.

Acknowledgments: To the Research Unit of Biotechnology and Genetic Resources germplasm bank, INIAV, Oeiras, Portugal, for providing the common bean samples. To FCT, Portugal, for the financial support in BEGEQA project (PTDC/AGR-TEC/3555/2012), E.M. PhD fellowship (SFRH/BD/89287/2012), as well as to R&D unit, UIDB/04551/2020 (GREEN-IT – Bioresources for sustainability). The authors acknowledge to the University of Düsseldorf for the collaboration in the experimental work and to COST Action FA1403 (STSM-FA1403-290815-063873) for funding. The authors also acknowledge PORTUGAL 2020 to the Portuguese Mass Spectrometry Network, grant number LISBOA-01-0145-FEDER-402-022125. The project NETDIAMOND (SAICTPAC/0047/2015), financially supported by FEEI (Lisboa 2020 and FCT/POCI-01-0145-FEDER-016385), to the iNOVA4Health (UID/Multi/04462/2013), financially supported by FCT and co-funded by FEDER under the PT2020 Partnership Agreement, as well as to POCI-01-0145-FEDER-029702, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES.

Author Contribution: Elsa Mecha: Conceptualization, Methodology, Investigation, Software, Data curation, Formal analysis, Validation, Visualization, Writing- Original draft preparation, Writing- Review & Editing **Guillaume L. Erny:** Software, Data curation, Writing- Review & Editing **Ana C. L. Guerreiro:** Methodology, Investigation, Software **Rodrigo P. Feliciano:** Methodology, Software, Supervision **Inês Barbosa:** Investigation **Andreia Bento da Silva:** Investigation **Susana T. Leitão:** Investigation**,** Software **Maria Manuela Veloso:** Resources **Diego Rubiales:** Resources **Ana Rodriguez-Mateos:** Supervision **Maria Eduardo Figueira:** Conceptualization, Funding acquisition **Maria Carlota Vaz Patto:** Conceptualization, Supervision, Funding acquisition, Project administration, Resources, Data curation, Writing- Review & Editing **Maria Rosário Bronze:** Conceptualization, Supervision, Funding acquisition, Project administration, Resources, Data curation, Writing- Review & Editing

Conflicts of Interest: The authors declare no conflict of interest.

Table 1. Tentative identification of metabolites, with mass accuracy (Δ mass) lower than 1 ppm and *mzCloud* match score higher than 80.0, in common bean accessions using *Compound Discoverer* software. For more details about the references please consult the list provided in Supplementary material. Consult **Table S2** for information regarding the remaining annotated metabolites

Clasification		#	Tentative Identification	Formula	RT (min)	ESI	Observed mass [M-H]-	Theoretical Molecular weight	Experimental Molecular weight	Δ mass (ppm)	Match m/z Cloud	FISh Cov.	Spectrum MS^2 (matched fragments)	Database (m/z Cloud ID; CS ID)	Reference(Plant/ Food item)
Phenylpropanoids and polyketides	Flavonoids (Flavonoid-3-O-glycosides)	Cp15	Rutin	$C_{27}H_{30}O_{16}$	7.224	-	609.14600	610.15338	610.15323	0.26	90.7	22.86	63.02348; 65.00316; 71.0135; 83.01385 ($C_4H_3O_2$); 93.03432; 107.01379; 108.02193 ($C_6H_4O_2$); 109.02912; 119.05016; 121.02897; 125.02473 ($C_6H_5O_3$); 148.01617; 151.00362 ($C_7H_3O_4$); 163.00389; 165.01881; 177.01878; 178.99857; 185.06102; 187.04041; 199.03963; 211.04004; 226.02658; 227.03464; 243.02951; 245.04579 ($C_{13}H_{19}O_5$); 254.02158; 255.02982; 271.02478; 271.06067; 272.03287; 283.02481; 299.01941; 300.0275 ($C_{15}H_8O_7$); 301.03555 ($C_{15}H_9O_7$); 609.14624 ($C_{27}H_{29}O_{16}$)	m/z Cloud (28) CS (4444362)	(Dueñas et al., 2015) (Common bean)
		Cp18	Quercetin-3β-D-glucoside or isomer	$C_{21}H_{20}O_{12}$	7.388	+	465.10266	464.09548	464.09546	0.04	99.3	62.50	53.03882; 55.01797 (C_3H_3O); 57.03354 (C_3H_5O); 61.02851 ($C_2H_5O_2$); 65.03843; 68.99693 (C_3HO_2); 69.03353 (C_4H_5O); 71.04902; 73.02831 ($C_3H_5O_2$); 81.03361 (C_5H_5O); 83.01222; 83.36504; 84.26059; 85.02836 ($C_4H_5O_2$); 87.04388 ($C_4H_7O_2$); 91.03872 ($C_3H_7O_3$); 93.03338; 97.02830 ($C_5H_5O_2$); 99.04412 ($C_5H_7O_2$); 108.24567; 109.02860 ($C_6H_5O_2$); 111.00771 ($C_5H_3O_3$); 119.29704; 121.02814 ($C_7H_5O_2$); 123.50369; 127.03898 ($C_6H_7O_3$); 137.02318 ($C_7H_5O_3$); 145.04955 ($C_6H_9O_4$); 145.06482; 149.02348 ($C_8H_5O_3$); 153.01817 ($C_7H_5O_4$); 155.04997; 159.36479; 163.03905; 164.88551; 165.01794 ($C_8H_5O_4$); 173.05930; 183.02849; 183.04318; 187.03889; 190.98940; 195.02887 ($C_9H_7O_5$); 201.05424; 205.04939; 219.06479; 228.04182; 229.04921; 247.05968 ($C_{13}H_{11}O_5$); 257.04413; 274.04538; 285.03873 ($C_{15}H_9O_6$); 286.04498; 303.04953 ($C_{15}H_{11}O_7$); 304.05255; 360.00449	m/z Cloud (1472) CS (4444361)	(Dueñas et al., 2015) (Common bean)
		Cp21	Astragalin (Kaempferol-3-O-glucoside)	$C_{21}H_{20}O_{11}$	7.669	-	447.09302	448.10056	448.10031	0.55	82.7	17.11	63.02389; 65.00321; 83.0136; 91.01852; 93.03452; 107.01379; 108.02133; 109.02919; 117.03447 (C_8H_5O); 125.02403; 132.02153; 135.00864; 137.02403; 143.05026; 145.02919; 151.00366; 153.019; 154.04243; 155.04982; 157.0659; 159.04495; 163.00334; 164.01151; 165.01952; 167.0497; 169.06581; 171.04466; 174.03226; 178.99818; 182.03697; 183.04501; 185.02484; 185.06064; 187.0399; 189.05603; 190.99911; 195.04393; 197.06096; 199.04019; 200.04802; 201.0553; 210.03252; 211.03999; 212.04787; 213.01994; 213.05571; 214.02745; 215.035; 226.02658; 227.03491; 228.04253; 229.05031 ($C_{13}H_9O_4$); 239.03462; 240.04268; 241.05019; 243.02991; 255.02974 ($C_{14}H_7O_5$); 256.03751; 257.04535 ($C_{14}H_9O_5$); 267.02914; 269.04529 ($C_{15}H_9O_5$); 283.02496; 284.0325 ($C_{15}H_{18}O_6$); 285.04034 ($C_{15}H_9O_6$); 299.05554; 327.05203; 447.09299 ($C_{21}H_{19}O_{11}$)	m/z Cloud (8165) CS (4445311)	(Lin et al., 2008) (Common bean)
	Flavonoids (Flavanonols)	Cp22	Taxifolin	$C_{15}H_{12}O_7$	7.678	+	305.06552	304.05830	304.05811	0.63	98.1	69.70	111.0439 ($C_6H_7O_2$); 121.0285 ($C_7H_5O_2$); 123.0441 ($C_7H_7O_2$); 139.0394 ($C_7H_7O_3$); 149.0235 ($C_8H_5O_3$); 153.0183 ($C_7H_5O_4$); 161.0235 ($C_9H_5O_3$); 167.0339 ($C_8H_7O_4$); 185.0599; 195.0287 ($C_9H_7O_5$); 213.0548 ($C_{13}H_9O_3$); 231.0652 ($C_{13}H_{11}O_4$); 259.0600 ($C_{14}H_{11}O_5$); 287.0548 ($C_{15}H_{11}O_6$); 305.0652 ($C_{15}H_{13}O7$)	m/z Cloud (3490) CS (388626)	(Ombra et al., 2016) (Common bean)
	Flavonoids (Flavonols)	Cp46	Quercetin	$C_{15}H_{10}O_7$	8.761	-	301.03510	302.04265	302.04246	0.65	94.2	15.38	63.02383; 65.00312; 65.00629; 83.01382; 93.03444; 107.01377; 109.02908; 121.02931; 124.01637; 139.03981; 149.02422 ($C_8H_5O_3$); 151.00354 ($C_7H_3O_4$); 159.04453; 161.02376; 164.01181; 169.01451; 178.99828; 187.03966; 193.0139; 201.05533; 227.03516; 229.04999; 245.04446; 255.02936; 273.03983 ($C_{14}H_9O_6$); 301.0351 ($C_{15}H_9O_7$)	m/z Cloud (27) CS (12269344)	(López et al., 2013) (Common bean)
						+	303.04959	302.04265	302.04231	1.12	98.6	54.55	65.03845 (C_5H_5); 68.99716 (C_3HO_2); 93.03331 (C_6H_5O); 95.04880; 109.02841 ($C_6H_5O_2$); 111.00748 ($C_5H_3O_3$); 115.05424; 121.02830 ($C_7H_5O_2$); 123.04385 ($C_7H_7O_2$); 131.04883; 137.02330 ($C_7H_5O_3$); 145.06441; 149.02342 ($C_8H_5O_3$); 151.03865 ($C_8H_7O_3$); 153.01826 ($C_7H_5O_4$); 155.04915; 159.04407; 161.05884; 163.03882 ($C_9H_7O_3$); 165.01797 ($C_8H_5O_4$); 166.02570 ($C_8H_6O_4$); 173.05937; 179.03448; 183.02866; 183.04422; 187.04001; 191.03358; 195.02794; 201.05464 ($C_{12}H_9O_3$); 219.06552 ($C_{12}H_{11}O_4$); 228.04173 ($C_{13}H_8O_4$); 229.04933 ($C_{13}H_9O_4$); 247.05934; 257.05081 ($C_{14}H_{10}O_6$); 274.04700 ($C_{14}H_{10}O_6$); 285.03879 ($C_{15}H_9O_6$); 303.04968 ($C_{15}H_{11}O_7$); 303.17813		

Table 1. *Cont.*

Classification		#	Tentative Identification	Formula	RT (min)	ESI	Observed mass [M-H]-	Theoretical Molecular weight	Experimental Molecular weight	Δ mass (ppm)	Match m/z Cloud	FISh Cov.	Spectrum MS^2 (matched fragments)	Database (m/z Cloud ID; CS ID)	Reference (Plant/ Food item)
Phenylpropanoids and polyketides	Flavonoids (Flavonols)	Cp50	Kaempferol	$C_{15}H_{10}O_6$	9.404	+	287.05469	286.04774	286.04762	0.40	99.5	51.11	68.99722 (C_3HO_2); 79.05413; 105.03358 (C_7H_5O); 107.04905 (C_7H_7O); 109.02851 ($C_6H_5O_2$); 111.00774 ($C_5H_3O_3$); 121.02856 ($C_7H_5O_2$); 133.02863 ($C_8H_5O_2$); 137.02344 ($C_7H_5O_3$); 147.04431 ($C_9H_7O_2$); 153.01840; 157.06502; 160.97350; 161.97682; 165.01855 ($C_8H_5O_4$); 171.04436; 183.02901; 185.05959; 213.05467 ($C_{13}H_9O_3$); 231.06577 ($C_{13}H_{11}O_4$); 241.04932 ($C_{14}H_9O_4$); 258.05252 ($C_{14}H_{10}O_5$); 259.05954 ($C_{14}H_{11}O_5$); 287.05499 ($C_{15}H_{11}O_6$); 287.14081	m/z Cloud (966) CS (4444395)	(López et al., 2013) (Common bean)
	Isoflavonoids (Isoflavones)	Cp32	Genistein	$C_{15}H_{10}O_5$	8.098	-	269.04523	270.05282	270.05257	0.94	87.3	13.33	91.01887; 133.0294; 135.00899; 153.01956; 157.06618; 169.06619; 180.05835; 181.06633; 183.04524; 197.06108; 199.03879; 207.04538; 225.05507; 241.05103 ($C_{14}H_9O_4$); 269.04535 ($C_{15}H_9O_5$)	m/z Cloud (24) CS (4444448)	(López et al., 2013) (Common bean)
		Cp39	Daidzein	$C_{15}H_{10}O_4$	8.524	-	253.05034	254.05791	254.05766	0.96	95.7	33.33	91.01878 (C_6H_3O); 132.0215; 133.02974; 135.00883 ($C_7H_3O_3$); 135.04501; 195.04478 ($C_{13}H_7O_2$); 196.05251; 197.06056; 208.05275; 209.06039; 223.04062 ($C_{14}H_7O_3$); 224.04826; 225.05556; 252.04214; 253.0504 ($C_{15}H_9O_4$)	m/z Cloud (680) CS (4445025)	(López et al., 2013) (Common bean)
		Cp44	Glycitein	$C_{16}H_{12}O_5$	8.633	-	283.06097	284.06847	284.06824	0.81	80.4	4.55	91.01885; 108.02149; 132.02179; 135.00879; 148.0166; 153.01938; 156.05812; 160.01695; 183.04543; 184.05299; 195.04518; 196.05325; 211.04012; 212.04794; 223.04013; 224.04764; 239.03484; 240.04282; 251.03471; 267.02979; 268.03754; 283.0611 ($C_{16}H_{11}O_5$)	m/z Cloud (428)	(Yang, Gan, Ge, Zhang, & Corke, 2018) (common beans)
	Aurone flavonoids (Auronols)	Cp34	Maesopsin (2,4,6-Trihydroxy-2-(4-hydroxybenzyl)-1-benzofuran-3(2H)-one)	$C_{15}H_{12}O_6$	8.206	-	287.05591	288.06339	288.06319	0.69	94.1	2.78	57.03433; 63.02388; 65.0032; 81.03452; 83.01379; 93.034;107.0138; 107.04974; 109.02916; 121.02886; 123.04467; 124.01658; 125.0243; 131.0499; 133.0302; 134.03682; 135.04494; 151.00362; 152.01143; 153.01945; 156.05838; 157.06549; 159.04526; 172.05269; 173.06056; 177.05565; 178.99837; 199.07549; 201.05568; 213.05539; 215.07123; 241.05026; 243.06662; 259.06094; 269.04568; 287.056 ($C_{15}H_{11}O_6$)	m/z Cloud (7874) CS (141288)	(Thuy et al., 2016) (*Artocarpus tonkinensis*)
Benzenoids	Benzene and substituted derivatives M-methoxybenzoic acids derivatives	Cp6	Vanillic acid	$C_8H_8O_4$	6.379	-	167.03491	168.04226	168.04216	0.57	87.3	50.00	108.02164 ($C_6H_4O_2$); 123.0444; 152.01141 ($C_7H_4O_4$); 167.03424	m/z Cloud (1471) CS (8155)	(Díaz-Batalla, Widholm, Fahey, Castaño-Tostado, & Paredes-López, 2006) (Common bean)
	Benzene and substituted derivatives (Salycillic acid)	Cp42	Salicylic acid	$C_7H_6O_3$	8.589	-	137.02420	138.03169	138.03158	0.84	98.5	85.71	65.0396 (C_5H_5); 93.03452 (C_6H_5O); 136.86255; 137.02434 ($C_7H_5O_3$)	m/z Cloud (643) CS (331)	(Radwan, Lu, Fayez, & Mahmoud, 2008) (*Vicia faba*)

Table 1. Cont.

Classification		#	Tentative Identification	Formula	RT (min)	ESI	Observed mass [M-H]-	Theoretical Molecular weight	Experimental Molecular weight	Δ mass (ppm)	Match m/z Cloud	FISh Cov.	Spectrum MS^2 (matched fragments)	Database (m/z Cloud ID; CS ID)	Reference (Plant/ Food item)
Lipids and lipid-like molecules	Fatty acyls (Medium chain fatty acids)	Cp30	Suberic acid	$C_8H_{14}O_4$	7.452	-	173.08174	174.08921	174.08909	0.70	93.3	66.67	57.03433; 83.05014; 109.0658 (C_7H_9O); 111.0814 ($C_7H_{11}O$); 129.0919; 173.0817 ($C_8H_{13}O_4$)	m/z Cloud (1393) CS (10025)	(FooDB_1.0) FDB003340 (Food and Fabaceae plants)
		Cp31	2-hydroxycaproic acid	$C_6H_{12}O_3$	8.013	-	131.07123	132.07864	132.07855	0.68	87.0	50.00	68.99545; 85.0658 (C_5H_9O); 87.04478; 131.07123 ($C_6H_{11}O_3$)	m/z Cloud (153) CS (90191)	(FooDB_1.0) FDB022697 (Food)
		Cp33	Azelaic acid	$C_9H_{16}O_4$	8.118	-	187.09749	188.10486	188.10473	0.68	96.3	60.00	57.03428; 69.03455; 83.05016; 95.05 (C_6H_7O); 97.06575; 123.08144 ($C_8H_{11}O$); 125.097 ($C_8H_{13}O$); 143.10765 ($C_8H_{15}O_2$); 169.08716 ($C_9H_{13}O_3$); 187.09734 ($C_9H_{15}O_4$)	m/z Cloud (331) CS (2179)	(FooDB_1.0) FDB012192 (Food and Fabaceae plants)
	Fatty acyls (Long chain fatty acids)	Cp69	16-Hydroxyhexadecanoic acid	$C_{16}H_{32}O_3$	15.352	-	271.22766	272.23514	272.23489	0.93	91.2	55.56	116.92824; 223.20638; 225.22226 ($C_{15}H_{29}O$); 253.21724 ($C_{16}H_{29}O_2$); 271.22763 ($C_{16}H_{31}O_3$)	m/z Cloud (2551) CS (10034)	(HMDB) HMDB0006294 (ChEBI) ChEBI 55328 (Plants)
	Prenol lipids (Triterpene saponins)	Cp60	(3β,5ξ,9ξ,18ξ)-22-Hydroxyolean-12-en-3-yl 6-deoxy-α-L-mannopyranosyl-(1->2)hexopyranosyl-(1->2)-β-D-glucopyranosiduronic acid	$C_{48}H_{78}O_{17}$	10.430	-	925.51642	926.52390	926.52368	0.24	94.8	65.12	67.0188 (C_4H_3O); 68.99813 (C_3HO_2); 69.03454; 71.01375 ($C_3H_3O_2$); 72.99297 (C_2HO_3); 73.02941 ($C_3H_5O_2$); 75.00863 ($C_2H_3O_3$); 83.01382 ($C_4H_3O_2$); 85.0294 ($C_4H_5O_2$); 86.00087 ($C_3H_2O_3$); 87.00861 ($C_3H_3O_3$); 87.04495; 89.02431 ($C_3H_5O_3$); 95.01371; 97.02949 ($C_5H_5O_2$); 99.00867 ($C_4H_3O_3$); 99.04505 ($C_5H_7O_2$); 101.02426 ($C_4H_5O_3$); 103.03989 ($C_4H_7O_3$); 111.00871; 112.01671; 113.02428 ($C_5H_5O_3$); 115.0036 ($C_4H_3O_4$); 115.03993 ($C_5H_7O_3$); 119.03474; 125.02441 ($C_6H_5O_3$); 127.04015 ($C_6H_7O_3$); 131.03481; 139.00389 ($C_6H_3O_4$); 143.03481 ($C_6H_7O_4$); 145.05037 ($C_6H_9O_4$); 157.01384 ($C_6H_5O_5$); 161.04543 ($C_6H_9O_5$); 163.06107; 205.07155 ($C_8H_{13}O_6$); 423.33044; 439.35764; 509.3996; 599.39862; 833.39819; 879.41077; 907.50928; 925.51624 ($C_{48}H_{77}O_{17}$)	m/z Cloud (8183) CS (22913504)	-
	Prenol lipids (Triterpene saponins)	Cp67	Ursolic acid	$C_{30}H_{48}O_3$	14.767	-	455.35272	456.36035	456.35995	0.87	80.5	33.33	79.9569; 319.22989; 455.35263 ($C_{30}H_{47}O_3$)	m/z Cloud (771) CS (58472)	(Seo et al., 2018) (Fruits and vegetables)

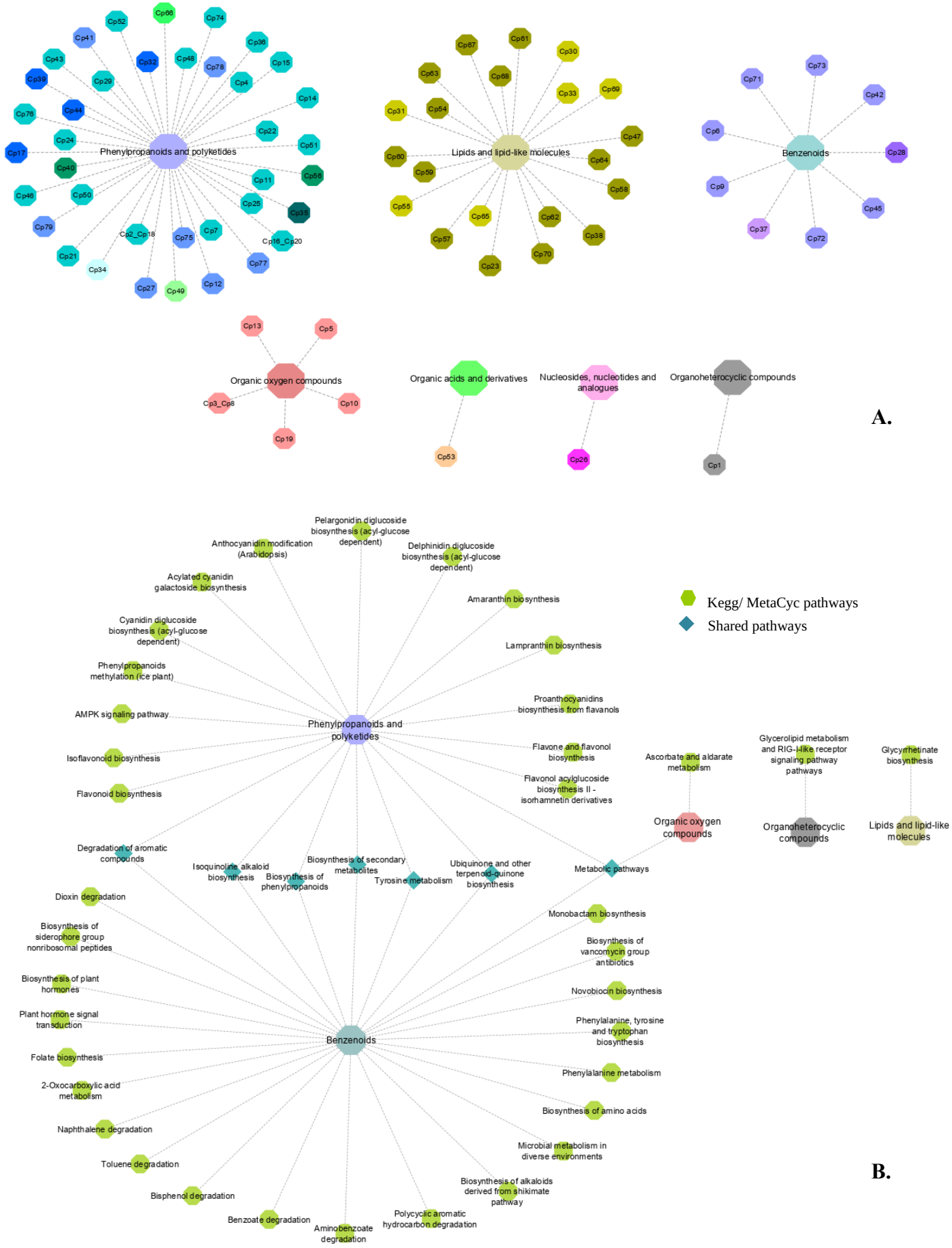

Figure 1. A. Classification of metabolites into different superclasses. **B.** Schematic representation of the described pathways in which the different superclasses of metabolites participate.

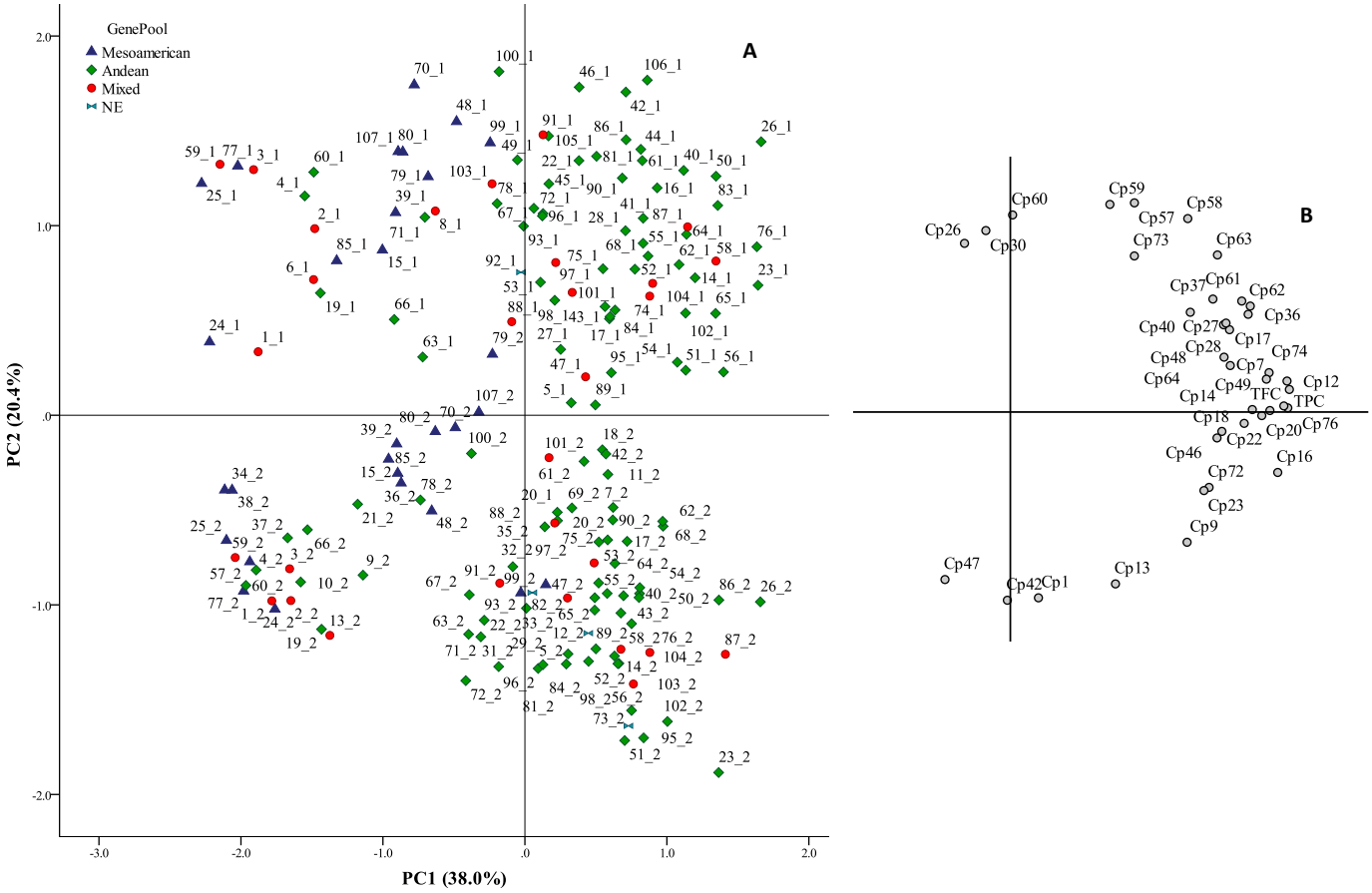

Figure 2. A. Score plot obtained by principal component analysis (PCA), showing common bean accessions, cropped in contrasting environments. The different accessions were named by the numbers attributed in **Table S1**, followed, after underscore, by the corresponding environment (1-Cabrela and 2-Córdoba). The gene pool of origin of the different accessions, Mesoamerican, Andean, mixed and NE – not evaluated, was superimposed in the representation using different symbols and colours. **B.** Correlation loading plot of parameters responsible by common bean accessions' projection, including TPC, total phenolic content; TFC, total flavonoid content; the area of compounds, named in accordance to **Table 1**, quantified by Q-Orbitrap-MS and the absolute concentration of compounds, named in accordance to **Table S5**, quantified by Q-TOF-MS.

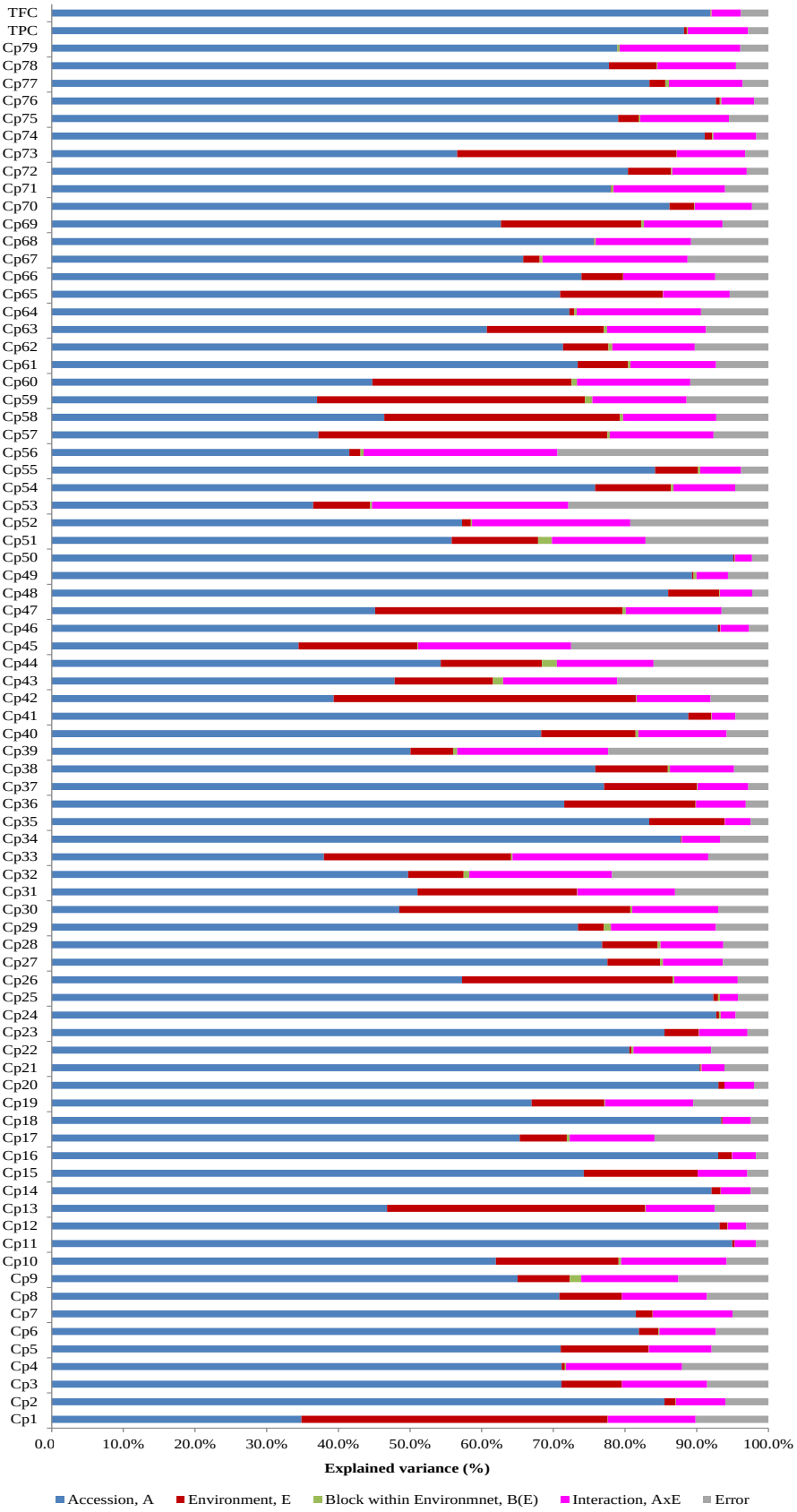

Figure 3. Contribution in % of **accession (A)**, **environment (E)**, **block within environment (B(E))**, and **accession x environment (AxE)** interaction to the variability of the analysed parameters.

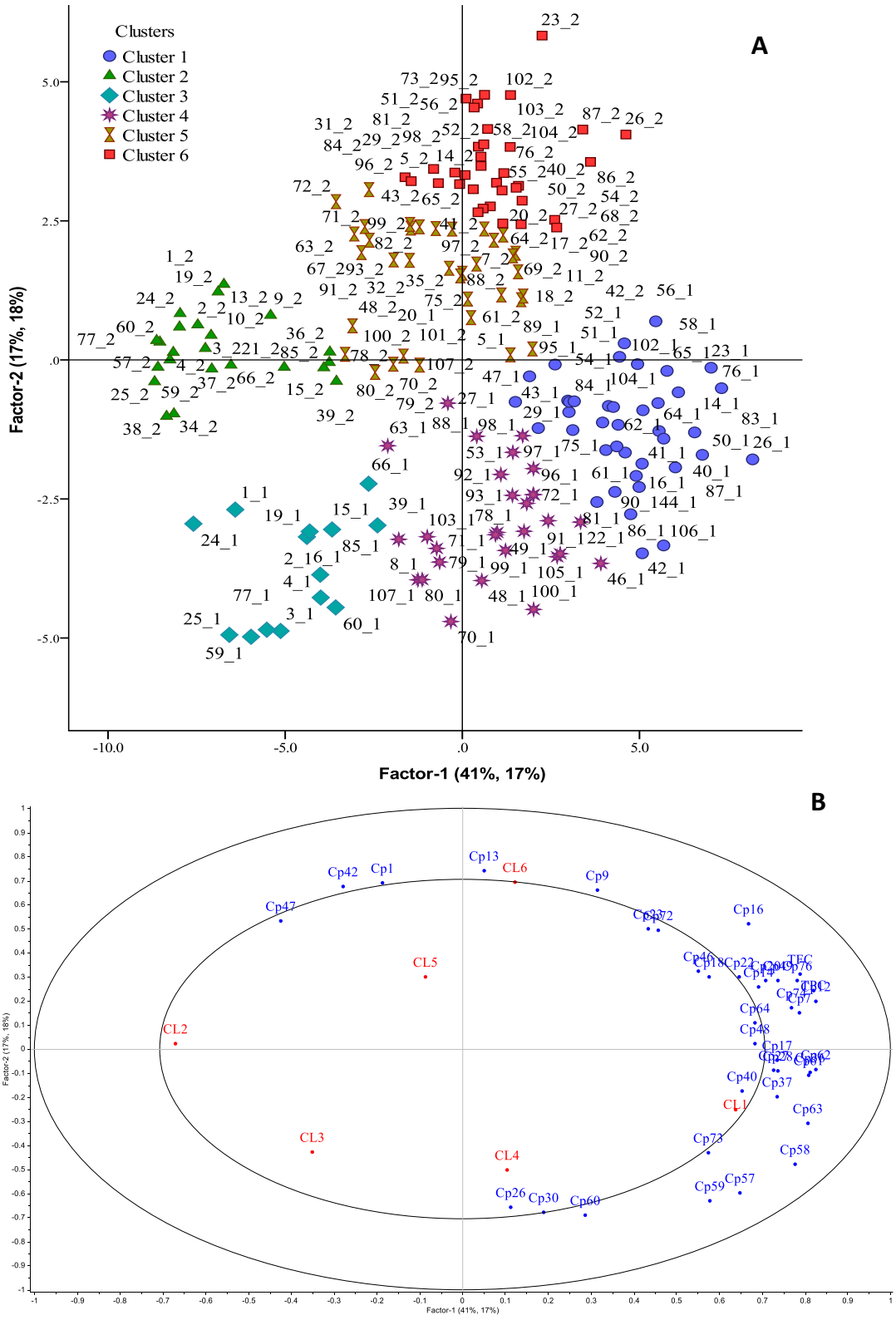

Figure 4. A. Score plot of common bean accessions, obtained by partial least square – discriminant analysis (PLS_DA), showing common bean accessions cropped in Cabrela and Córdoba, grouped into different clusters along the two first factors. The explained variance (%) of predictors (X variables) and responses (Y variables) attributed to the first and second component, factor # (X%, Y%), are shown in the figure. The different common bean accessions were named as reported in **Figure 2**. **B.** Correlation loading plot of clusters and parameters, named in accordance to **Figure 2**. The majority of parameters were located between the inner and outer (50 and 100%) explained circles. $R^2(X) = 0.5840$; $R^2(Y) = 0.3474$; RMSECV = 0.2988; RMSEC = 0.2934; $Q^2 = 0.1331$ and $R^2\text{-}Q^2 = 0.2143$, difference < 0.3 (Kiralj & Ferreira, 2009), indicate the quality of the model.

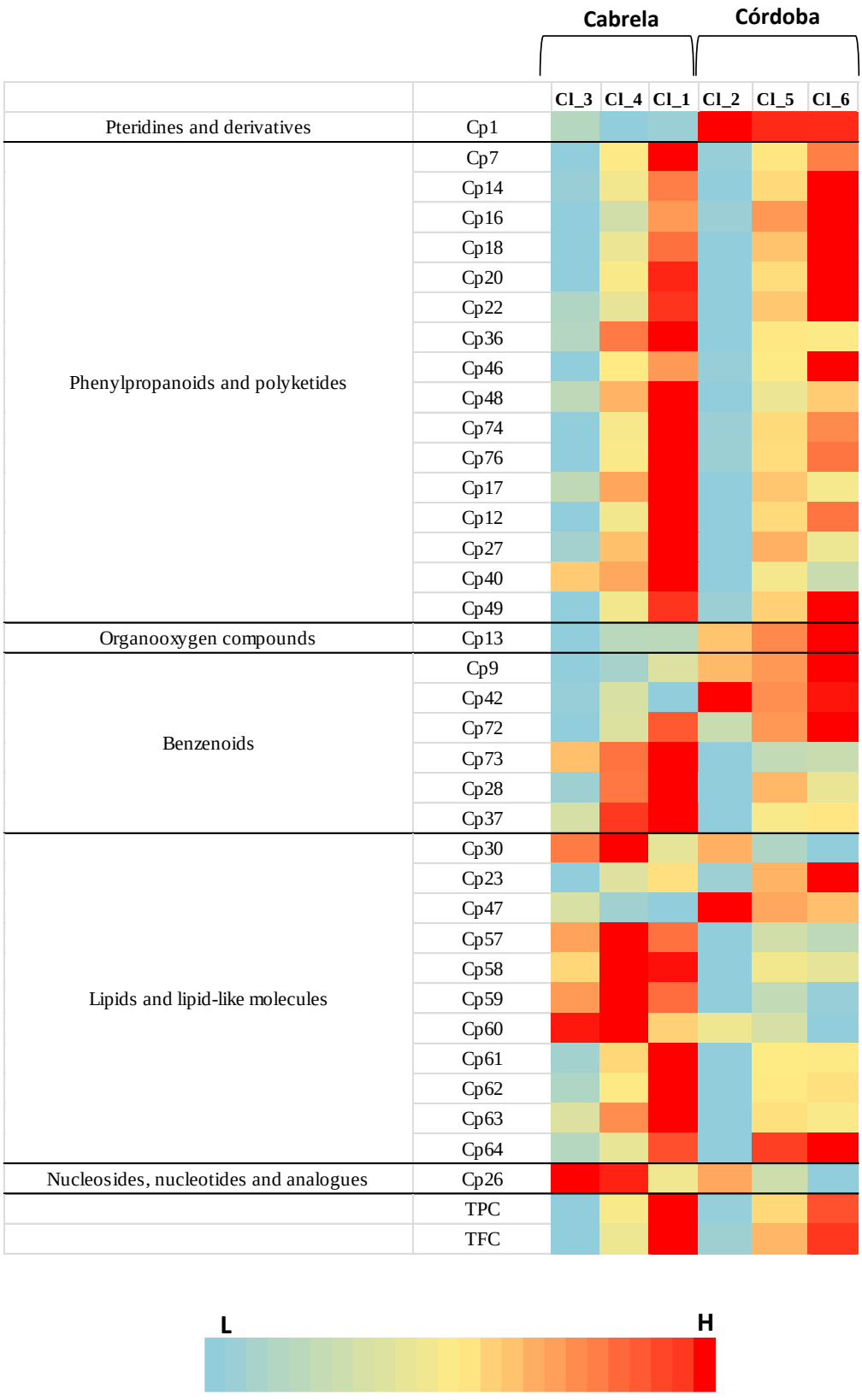

Figure 5. Heat map showing the abundance of the parameters (low - L to high - H) responsible by common bean accessions' projection, **Figure 4**, into the different clusters. Clusters 3, 4 and 1 were mostly characterized by common bean accessions from Cabrela environment and clusters 2, 5 and 6 included mostly accessions from Córdoba. For data details please consult **Table S7**.