Special topic paper

Greta Bindi, Andrew Smith*, Paola Crisafulli, Vanna Denti, Eleonora Bossi, Simone Serrao, Isabella Piga, William S. Sánchez, Giuseppe Paglia, Fulvio Magni and Luciano Navarini

Mapping the distribution of bioactive compounds and aroma/flavour precursors in green coffee beans with an integrated mass spectrometry-based approach

https://doi.org/10.1515/pac-2024-0210

Abstract: Coffee is one of the most enjoyed beverages worldwide and whilst a wide array of secondary metabolites has been qualitatively and quantitatively characterised within the coffee bean, little is known about their distribution in different bean tissues. In this work, complementary mass spectrometry (MS) based approaches were utilised to determine the spatial metabolomic content of coffee beans, focusing on highlighting those which presented a regiospecific distribution and distinguished the different endosperm regions and the embryo. Preliminary exploration of the spatial metabolome of *Coffea arabica* coffee beans highlighted tissue-specific molecular features through an unsupervised statistical analysis. Then, the complementary information derived from higher resolution MS approaches resulted in the annotation of 13 metabolites which showed a specific enrichment in at least one tissue, and most of these were noted to be bioactive compounds or aroma/flavour precursors. Knowledge of the tissue distribution of these compounds could shed light on their biological role in plant physiology and bean development, but could also have relevant implications for their positive impact on human health due to coffee consumption as well as their possible effect on aroma and flavour upon roasting of green coffee beans.

Keywords: Aroma precursors; bioactive compounds; coffee beans; flavour precursors; liquid chromatography; MASSA 2023; mass spectrometry; mass spectrometry imaging; spatial metabolomics.

Introduction

Thanks to its pleasant taste and aroma, as well as its functional properties deriving primarily from caffeine, coffee is one of the most enjoyed beverages worldwide [1]. The beverage is obtained by the roasting, grinding, and brewing of green coffee beans, which are the half-ellipsoidal seeds of plants belonging to the Rubiaceae family, genus *Coffea*, particularly *Coffea arabica* L. and *C. canephora* Pierre ex A. Froehner, the two species with

Article note: A collection of invited papers based on presentations at the Mass spectrometry congress in Italy – MASSA 2023 held on 28–30 June 2023 in Torino, Italy.

***Corresponding author: Andrew Smith**, Department of Medicine and Surgery, Proteomics and Metabolomics Unit, University of Milan-Bicocca, Edificio U28, Via Raoul Follereau, 3, Vedano al Lambro (MB), 20854, Monza, Italy, Phone: +39 0264488204, e-mail: andrew.smith@unimib.it
Greta Bindi, Vanna Denti, Eleonora Bossi, Simone Serrao, Giuseppe Paglia and Fulvio Magni, Department of Medicine and Surgery, Proteomics and Metabolomics Unit, University of Milano-Bicocca, Monza, Italy
Paola Crisafulli and Luciano Navarini, illycaffè S.p.A., via Flavia 110, Trieste I-34147, Italy
Isabella Piga, Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy
William S. Sánchez, Agriculture, Livestock and Agroforestry Program, CATIE, Turrialba, Costa Rica

commercial interest among more than 100 species not yet industrially exploited [2, 3]. Coffee beans are mainly composed by a reserve tissue known as the endosperm and surrounded by a silverskin, a thin fibrous layer which is a remnant of the perisperm. The endosperm includes an epidermal layer composed by a single cell layer, a hard external region composed by poligonal cells, and a soft internal region composed by rectangular cells that surround the embryo chamber [4–6]. The endosperm cells supply nutrients to the embryo that develops inside the embryo chamber until the germination step [3, 4, 7]. The embryo is a 3–4 mm structure composed of a hypocotyl axis and two cotyledons, generally localised near the convex, external portion of the seed [4, 7].

In addition to caffeine, green coffee is characterised by the presence of numerous secondary metabolites which contribute to the development of its well-known flavour upon roasting, as well as to exert bioactivity, promoting beneficial effects on human health [8]. For instance, caffeine and other minor methylxanthines in coffee, such as theophylline and theobromine, as well as the furokaurane glucoside mozambioside, are well-known bitter-tasting compounds [9–11]. Further examples may be represented by chlorogenic acids and trigonelline. Chlorogenic acids (CGAs) are a large class of esters formed between quinic acid and hydroxycinnamic acids [12]. They are present in coffee as a complex mixture of positional and geometric isomers, where caffeoylquinic acids (CQAs) are the most abundant, followed by di-caffeoylquinic acids (diCQAs), feruloylquinic acids (FQAs), p-coumaroylquinic acids (p-CoQAs), and many others in trace amounts [13, 14]. These coffee polyphenols have been suggested to be involved in the plant defence mechanism against environmental aggressions and, upon roasting, are partially converted into phenolic lactones known as quinides, which contribute to the bitter taste of the beverage [15], as well as into aroma compounds such as guaiacol, catechol, and their derivatives, remarkably important in providing relevant sensory notes to the brews [16]. Moreover, CGAs have recently attracted the attention of several research groups thanks to their biological activity including antioxidant, anti-inflammatory, anti-diabetic, anti-carcinogenic, and antiviral properties [17–20]. After caffeine, trigonelline (1-N-methylnicotinic acid, nicotinic acid N-methylbetaine) is the second most abundant alkaloid compound in raw coffee beans [21] and, upon bean roasting, it is partially decomposed to give volatile degradation products such as pyridines, very important coffee aroma compounds, as well as non-volatile bioactive compounds such as N-methylpyridinium (NMP) ions *via* decarboxylation as well as nicotinic acid and nicotinamide (both exerting vitamin B3 activity) *via* demethylation [22]. The possible effects of trigonelline on health mediated *via* hypoglycaemic, neuroprotective, anticancer, estrogenic, and antibacterial activities have been reported in literature [23–25]. Some other bioactive compounds found in coffee, such as isoflavones [26] and flavonols [27], have not yet been studied as possible aroma/taste precursors as far as we know, whereas other compounds, including sucrose and aminoacids, are well known and important flavour precursors [28].

These coffee phyto-chemicals can be quali-quantitatively characterised by resorting to several analytical techniques including gas chromatography (GC) and liquid chromatography (LC), often coupled to mass spectrometry (MS) as well as ion mobility separation (IMS). Moreover, the advent of high-resolution mass spectrometry (HRMS) techniques has further facilitated the extensive detection and identification of coffee compounds [29, 30]. Nevertheless, whilst traditional bulk approaches are useful for the identification and the quantification of a wide range of coffee secondary metabolites, these techniques fail to explore their tissue distribution, precluding the possibility to visualise which tissues of the bean are most enriched in some analytes of interest. In fact, on the one hand, understanding the accumulation and distribution of different molecules in tissue could shed light regarding their plant physiological function and, on the other hand, the possibility to determine which regions of coffee beans are mostly enriched in bioactive compounds could potentially have a positive impact in terms of better understanding the effect of the roasting process on these compounds, and then on human health upon coffee consumption [1, 3]. Mass Spectrometry Imaging (MSI) is a powerful technique that combines the ability to investigate the content of samples in an untargeted manner, with the possibility to resolve the tissue distribution of various classes of biomolecules [31], and recently it is being increasingly employed to map the spatial distribution of endogenous low molecular weight molecules in plants, including coffee beans [30, 32–36].

In this context, this study aims to utilise Matrix-Assisted Laser Desorption/Ionisation – Time of Flight (MALDI-TOF) MSI to assess which molecules best characterise the epidermal layer, the hard endosperm, the soft endosperm and the yet unexplored embryo region of coffee beans. To achieve this aim, *C. arabica* has been considered in view of its higher value in the coffee market thanks to its smooth, mild, and rich flavour [37]. The

whole set of experimental data has been compared to that obtained from *Coffea pseudozanguebariae* sample. This caffeine-free wild coffee species [38] has been used as a control in consideration of its very low content of chlorogenic acids compared with other wild coffee species [39] and in view of the high mozambioside content compared to that of *C. arabica* L. [40]. Moreover, MALDI-MSI data was subsequently correlated with complementary mass spectrometry approaches, namely ion mobility separation (IMS) MALDI-qTOF-MSI and Ultra-High Performance LC (UHPLC) coupled to ESI-MS/MS, in order to identify more extensively those analytes that present a specific bean distribution, focusing on bioactive compounds and flavour precursors. Collectively, this highlighted those principal bioactive compounds which present a tissue specific distribution within the coffee bean, suggesting their possible role in plant physiology. Moreover, similar findings were also observed for well-known flavour precursors, such as sucrose, amino acids, trigonelline and chlorogenic acids, indicating their possible exposure to thermal degradation induced by the roasting process of green coffee beans. In particular, the observed loss of chlorogenic acid during roasting (up to 95 % in dark roasted coffee), which has been extensively reported [41, 42], is a consequence of the thermal degradation, which starts with isomerization and epimerization processes in the initial roasting stages, followed by lactonization and degradation reactions in the later stages. Moreover, sucrose trigonelline and amino acids rapidly degrade at the early stage of roasting and instantly participate in manifold reactions. Sucrose fast hydrolysis at the beginning of the roasting process releases the reducing sugars glucose and fructose, which thereupon are strongly involved in both caramelization and together with amino acids Maillard-type reactions able to generate the important volatile organic compounds characterizing roasted coffee aroma. It is clear that the spatial distribution of aroma/flavour precursors within the bean can influence their susceptibility to thermal degradation and, in turn, can direct towards the optimization of the roasting process.

Materials and methods

Chemicals and reagents

HPLC-grade water, HPLC-grade acetonitrile, HPLC-grade methanol, formic acid, and ammonium formate were obtained from Honeywell SC, Seelze, Germany. Trifluoracetic-acid (TFA) and phosphorus red were obtained from Sigma-Aldrich, Buchs, Switzerland. 2,5-dihydroxybenzoic acid matrix was purchased from Bruker Daltonics, Bremen, Germany. MMI-L Low Concentration Tuning Mix was obtained from Agilent Technologies, Santa Clara, CA, USA.

Coffee samples

In this study, a sample of wet-processed green coffee beans (*C. arabica* L.) from Ethiopia (Crop 2019–2020) was used. The sample with zero primary and secondary defects was selected on the basis of standard procedures of sorting and visual aspect, moisture content, screen size, and cup quality by illycaffè S.p.A. Quality Control Dept. (Trieste, Italy). A sample of green coffee beans in parchment (*C. pseudozanguebariae* Bridson) from CATIE (Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica) germplasm (code T.21352) was also used. In a previous study, the latter genotype was found to be the coffee species with the lowest total CGAs content among ten different coffee species including *C. arabica* [43].

Sample preparation for MALDI-MSI

For each coffee species, four ten-micron-thick sections containing the embryo of the bean were cryosectioned and mounted onto conductive Indium Tin Oxide (ITO) slides (Bruker Daltonics, Bremen, Germany). Matrix deposition for MALDI-MSI analysis was carried out by spraying 2,5-dihydroxybenzoic acid (40 mg/ml in 70 % MeOH, 0.1 %

TFA) using a HTX TM-Sprayer™ (HTX Technologies, LLC, Chapel Hill, NC, USA) with the following parameters: temperature 80 °C; number of passes 5; flow rate 0.13 ml/min; velocity 1200 mm/min; track spacing 2 mm; pressure 10 psi.

Sample preparation for UHPLC-ESI-IMS-MS/MS

C. arabica and *C. pseudozanguebariae* green coffee beans were mechanically ground, and the obtained powder was placed in an Eppendorf tube (Eppendorf, Hamburg, Germany) with 100 % MeOH to reach a final concentration of 10 mg/ml and centrifuged for 1 min at 13,000 rpm; the supernatant was then transferred into a clean Eppendorf tube for UHPCL-ESI-MS/MS analysis.

MALDI-MSI analysis

A rapifleX MALDI Tissuetyper™ (Bruker Daltonics GmbH, Bremen, Germany) was used to analyse one section per bean in reflectron positive ion mode, and one in negative ion mode, within the *m/z* 100–1000 range. MALDI-TOF MSI images were acquired with a beam scan setting of 46 µm and a raster sampling of 50 µm in both *x* and *y* dimensions. Red phosphorus was directly applied onto the ITO glass slide to be used as an external calibration standard, approximately within the *m/z* 150–1000 range. FlexControl 4.0 (Bruker Daltonics, Bremen, Germany) was used to set up instrument parameters for the acquisition method and FlexImaging 5.0 (Bruker Daltonics, Bremen, Germany) for the visualisation of MALDI-MSI analyses.

Following the MALDI-TOF MSI analysis, the sections of *C. arabica* that were analysed in positive and negative ion mode were removed from the ITO slide and placed onto a clean glass microscope slide. Subsequently, the sections were washed with ethanol (100 % × 30 s, 95 % × 30 s) to remove the residual DHB matrix and were rehydrated in MilliQ water for 2 min. Slides were stained with an aqueous Toluidine Blue O solution (TBO 0.1 % in MilliQ water for 30 s) and rinsed with water for 3 min [44]. Lastly, slides were dried and mounted with permanent mounting medium. Following histological staining, slides were scanned using a digital scanner (NanoZoomer, Hamamatsu Photonics) and converted to digital format.

MALDI-IMS-qTOF-MSI analysis was performed on two coffee sections for each polarity using a timsTOF fleX (Bruker Daltonics GmbH, Bremen, Germany) to profile a portion of each bean with a beam scan setting of 46 µm and a raster sampling of 50 µm in both *x* and *y* dimensions; red phosphorus was used as an external calibrant. Tune settings are reported in Supplementary Information, S1.

UHPLC-ESI-IMS-MS/MS analysis

The extracted metabolites were analysed using an UHPLC-ESI-IMS-MS/MS platform which includes an Elute UHPLC system coupled to a timsTOF flex spectrometer. RP chromatographic separation was achieved using an InfinityLab Poroshell 120 EC-18 colomn 3 × 150 mm 2.7 µm (Agilent Technologies) and an ACQUITY UPLC BEH Amide column, 130 Å, 1.7 µm, 2.11 mm × 150 mm (Waters Corporation) was used to separate more polar compounds. Chromatographic conditions and mass spectrometry analysis parameters are reported in Supplementary Information, S2.

MALDI-MSI data analysis

After the MALDI-TOF-MSI analysis, SCiLS Lab 2023a Pro software (http://scils.de/; Bremen, Germany) was used to generate two files containing respectively measurements obtained in positive and negative ion mode. Data was pre-processed by performing baseline subtraction (Convolution algorithm), normalisation (Total Ion Current

algorithm) and weak spatial denoising. Peak picking was performed to eliminate chemical noise and generate a *m/z* feature list. Individual regions were then subjected to automatic segmentation using the k-means algorithm to cluster pixels based on their spectral similarity; weak denoising was applied to account for the intervariability of adjacent pixels. Subsequently, histological images were imported in the SCiLS Lab files and co-registered with the respective molecular images in positive and negative ion mode to integrate metabolomic and morphological data. Histological images were used as a reference to draw regions of interest onto the imaging dataset and obtain the mean spectra of the principal tissues of the coffee bean: the mean mass spectra of these regions of interest (ROIs) were calculated and used to perform Principal Component Analysis (PCA) to emphasise the differences amid ROIs while reducing the complexity of the dataset. Receiver-operating characteristic (ROC) analysis was performed to compare the intensity of peaks among the embryo, the endosperm and the epidermal layer. Lastly, the mean intensities of features of interest within the embryo, epidermal layer and endosperm of *C. arabica* and *C. pseudozanguebariae* coffee bean sections were obtained and imported in Metaboanalyst 5.0, an open-source omics data analysis platform, to generate a heatmap using the Euclidean distance measure and Ward clustering method as parameters.

Data derived from the MALDI-IMS-qTOF MSI analysis was imported in MetaboScape® (Bruker Daltonics GmbH, Bremen, Germany) to perform tentative annotation of small molecules. Annotation was performed setting a tolerance of 20.0 ppm for Δ*m/z* and 15.0 for ΔCCS% score, considering $[M+H]^+$ as primary ion and $[M+Na]^+$ and $[M+K]^+$ as seed ions in positive ion mode, and $[M-H]^-$ as primary ion and $[M+Cl]^-$ as seed ion in negative ion mode. When CCS values were not present in libraries, CCS-Predict Pro algorithm was used to predict a theoretical CCS value based on InChi code structures. CCS-Predict Pro was not available for ions $[M+K]^+$ and $[M+Cl]^-$, and thus the theoretical CCS value of these molecular species could not be obtained.

Lastly, the MALDI-TOF MSI mean spectra obtained in positive and negative ion mode were imported in mMass, an open source mass spectrometry tool, to perform external calibration using 7 reference masses of known metabolites that were present in the UHPLC-ESI-IMS-MS/MS dataset. Tentatively annotated compounds from the MALDI-IMS-qTOF-MSI dataset were then correlated with molecular features that presented a specific distribution within the bean, emerged through the ROC analysis, from the MALDI-TOF MSI dataset by setting a mass error tolerance of 0.06 Da, as ±0.06 was designated as the average two-sided uncertainty to be added around the central *m/z* value to best describe the width of peaks in this dataset.

UHPLC-ESI-IMS-MS/MS data analysis

UHPLC-ESI-IMS-MS/MS data was imported in Metaboscape® (Bruker Daltonics GmbH, Bremen, Germany) to perform tentative annotation of features based on mass error (M_e < 10 ppm), isotopic pattern (mSigma < 200), MS/MS fragmentation pattern (MSMS score >200, when available) and ion mobility (ΔCCS% < 15.0), considering $[M+H]^+$ as primary ion in positive ion mode, and $[M-H]^-$ as primary ion in negative ion mode. When CCS values were not present in libraries, CCS-Predict Pro algorithm was used to predict a theoretical CCS value based on InChi code structures. CCS-Predict Pro was not available for ions $[M+K]^+$ and $[M+Cl]^-$, and thus the theoretical CCS value of these molecular species could not be obtained.

Results

MALDI-MSI analysis of coffee beans

In this pilot study, *C. arabica* green coffee beans were analysed with MALDI Mass Spectrometry Imaging to highlight the possibility to explore the metabolomic content of different tissues within the bean, including the epidermal layer, that is the outer monolayer of the endosperm, the hard endosperm region, the soft endosperm region and the embryo. Figure 1a displays the results of the histological staining of a 10 µm thick coffee section with a toluidine blue solution, which was used as a reference to manually annotate the main tissues of the bean

(Fig. 1b). Semi-supervised Principal Component Analysis was performed to determine whether there were metabolomic differences among the various tissues of the bean, and the scores plot was able to identify separate clusters representing the embryo (yellow), the epidermal layer (red), and a third cluster composed of spectra derived from both the soft and hard endosperm (blue and green). Interestingly, the embryo emerged as the most distinct region, considering that the distance of clusters correlates to the differences in terms of metabolomic content and the most distant cluster is represented by the spectra of the embryo region, as shown in Fig. 1c [45]. For analogous reasons, the high degree of overlap observed for the soft and hard endosperm clusters indicates a prominent similarity in terms of their metabolomic content. The peculiarities of the metabolomic profiles of the different regions of the coffee bean were also evident in the generated mean spectra (Fig. 1d), where, interestingly, the embryo appeared to be characterised by a greater number of unique features and a generally higher intensity of peaks, especially within the *m/z* 400–700 mass range of the negative ion mode dataset.

Based on these preliminary results, spatial segmentation with the k-means clustering algorithm was performed to reduce the dimensionality of the dataset by identifying regions with a similar metabolic landscape,

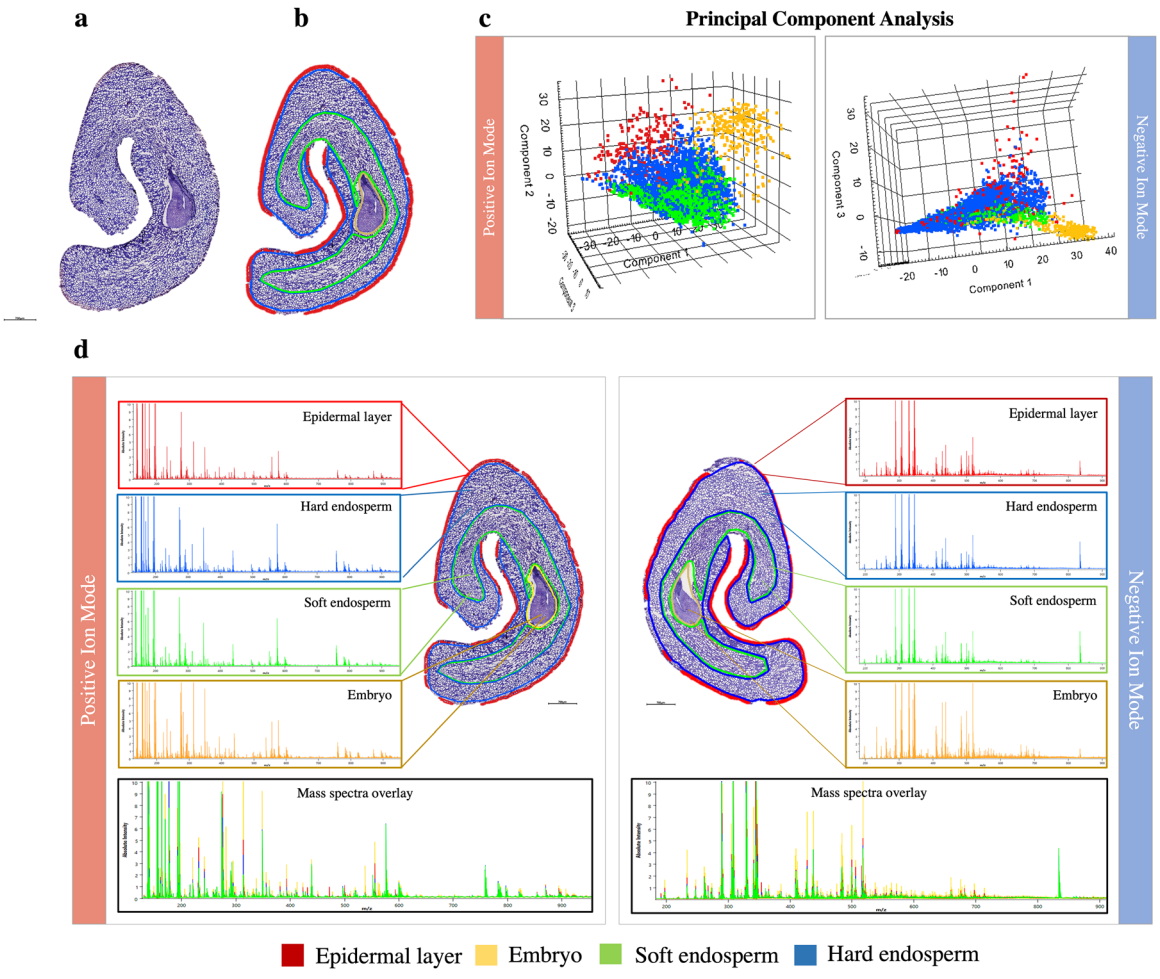

Fig. 1: a. TBO stained *C. arabica* bean; b. manual annotations of histological image highlighting the main tissues within the bean: the epidermal layer (red), the hard endosperm (blue), the soft endosperm (green) and the embryo region (yellow); c. semi-supervised principal component analysis performed on positive and negative ion mode datasets allows to separate the main tissues of the bean according to their metabolomic content, with the embryo region showing the most diversity; d. comparison of the mean mass spectra of the main tissues of the bean allows to identify unique peaks that show a differential distribution among tissues.

clustering pixels based on their spectral similarity [45]. Spatial segmentation on samples analysed in positive and negative ion mode shed light on a distinct molecular content of the bean, which was organized in 12 and 25 clusters, respectively, that were similar in terms of metabolomic profile. Consequently, it was also possible to confront the known tissues of the beans highlighted by the histological staining with the clusters that emerged through the unsupervised approach, as shown in Fig. 2. Specifically, spatial segmentation could efficiently distinguish a cluster, highlighted in dark blue in both datasets, which showed a strong co-localization with the embryo region. Interestingly, there weren't any clusters that specifically co-localized with the soft or hard endosperm of the bean. Lastly, while the histological staining highlighted a monolayer of cells that could be annotated as epidermal layer, the unsupervised clustering underlined four concentric clusters in positive ion mode (dark blue, blue, light blue and purple) and two clusters in negative ion mode (blue and purple) which were localised towards the external, convex portion of the bean, and, although partially overlapped, altogether regarded a larger area compared to the epidermal layer.

To determine which metabolite features were primarily responsible for the differences among the bean tissues, as highlighted by the PCA and spatial segmentation, ROC analysis was performed; considering that both PCA and spatial segmentation indicated a high degree of metabolomics similarity among the soft and hard endosperm, these tissues were unified and considered as a single region in the following comparisons. In the positive ion mode dataset, 70 features were determined to be increased in the embryo (AUC > 0.75), 3 in the endosperm (AUC > 0.75), and 5 in the epidermal layer (AUC > 0.75); regarding the negative ion mode dataset, 88 features were determined to be increased in the embryo (AUC > 0.75), 1 in the epidermal layer (AUC > 0.75), and 0 in the endosperm (AUC > 0.75).

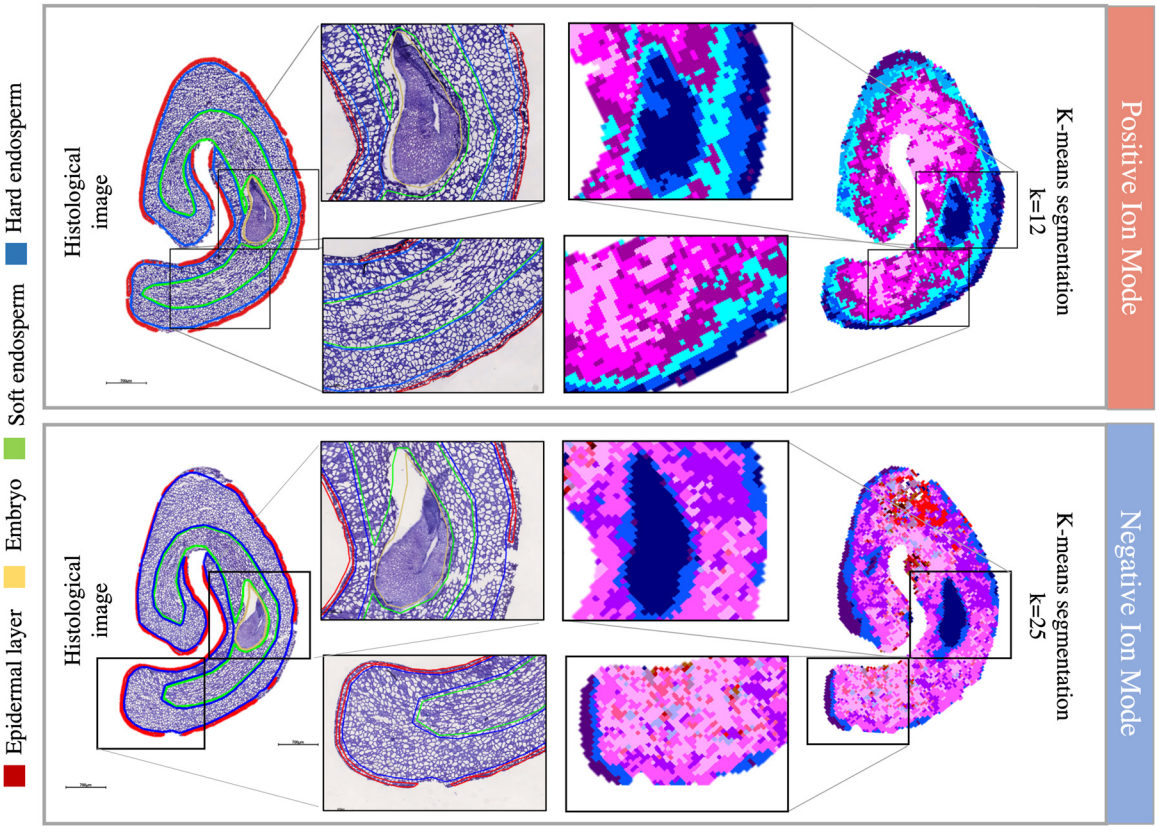

Fig. 2: Comparison of TBO stained *C. arabica* sections and spatially segmented MALDI-MSI with k-means algorithm in positive and negative ion mode. Each cluster defined by the k-means algorithm is represented by an arbitrary colour.

MALDI-IMS-MSI profiling of coffee beans

To annotate features highlighted by the ROC analysis in the MALDI-TOF MSI dataset, MALDI-qTOF Ion Mobility Separation MSI (IMS-MSI) was performed on *C. arabica* sections to profile and putatively annotate features based on their *m/z* ratio and Collision Cross Section (CCS) value. Putative annotations comprised a variety of metabolites including alkaloids such as caffeine and theobromine, isoflavones and diterpenes; furthermore, a rich lipidome of triacylglycerols and phospholipids was indicated. Interestingly, putative annotations did not only comprise the most characterised chlorogenic acids, including caffeoylquinic and di-caffeoylquinic acids, but also those that are less abundant, such as p-coumaroylquinic and sinapoylquinic acids, which are known to be present in *C. arabica* beans, although in low amount [14]. Of these putative annotations, 29 could be correlated with the MALDI-TOF-MSI datasets when setting a mass error tolerance of 0.06 Da, thus enabling these annotated compounds to be associated with a specific distribution within the coffee bean (Table 1). For two of these compounds, caffeine and sucrose, it was also possible to obtain an *in situ* MS/MS spectrum including fragment ions which are reported to be diagnostic for these metabolites [46, 47], which validated putative annotations obtained with MALDI-qTOF-IMS, as shown in Supplementary Information, S3.

UHPLC-ESI-IMS-MS/MS analysis of coffee beans

Lastly, UHPLC-ESI-IMS-MS/MS was employed to identify compounds within *C. arabica* coffee bean samples with a high-resolution mass spectrometry approach, adding a further confidence level to those already annotated through MALDI-IMS-qTOF-MSI. 13 compounds annotated with UHPLC-ESI-IMS-MS/MS were also present within the list of metabolites of interest emerged through the correlation of MALDI-TOF MSI and MALDI-IMS-qTOF-MSI data (Table 1), as shown in Table 2. The annotated compounds comprised alkaloids (trigonelline, theophylline, caffeine), an amino acid (arginine), a disaccharide (sucrose), the most abundant chlorogenic acids (CQA regioisomers, di-CQA regioisomers), and a CGA derivate (di-CQL regioisomers), an isoflavone (daidzein), a diterpene (cafestol) and phospholipids (LPI 16:0, LPI 18:2, PC 34:2).

Integrated data

Having identified a list of 13 putative annotations that were present both in the UHPLC-ESI-IMS-MS/MS and MALDI-IMS-qTOF-MSI datasets, and could be successfully correlated with 15 features present in the MALDI-TOF MSI dataset, it was possible to highlight the specific molecular distribution of these identified compounds and underline their varied abundance in either the embryo, the endosperm, or the epidermal layer (AUC > 0.75), as shown in Fig. 3.

Lastly, to explore the specificity of these distributions in *C. arabica* coffee beans, the mean intensities of features of interest within the different bean tissues were obtained and confronted with the same intensities within a single section of *C. pseudozanguebariae* obtained with MALDI-TOF MSI; the species *C. pseudozanguebariae* was purposely selected due to its acknowledged deficiency of caffeine and low CGAs content, so that the ratio of intensities of these metabolites within different bean tissues in this specimen could shed further light regarding their physiological function in the coffee bean [48]. The obtained intensities were then used to generate a heatmap for every bean, as shown in Fig. 4. When visually confronting the heatmaps representing the intensities of metabolites within *C. arabica* and *Coffea pseudozaguebariae* coffee bean sections, some displayed a similar trend amongst the two specimen, including Caffeine $[M+H]^+$, Arginine $[M+H]^+$, Daidzein $[M+Cl]^-$, CQA regioisomers $[M-H]^-$, di-CQL regioisomers $[M+Cl]^-$, di-CQA regioisomers $[M+Cl]^-$, LPI 16:0 $[M-H]^-$, and LPI 18:0 $[M-H]^-$. Instead both adducts of Trigonelline ($[M+Na]^+$, $[M+K]^+$), but also Theophylline/Theobromine $[M+H]^+$, Cafestol $[M+K]^+$, Sucrose $[M+K]^+$, and CQA regioisomers $[M+K]^+$ resulted enriched in the embryo of the species *C. Arabica*, whilst their abundance was increased in the epidermal layer of *C. pseudozanguebariae*. Lastly, while PC 34:2 $[M+H]^+$ displayed a clear enrichment in the endosperm of the

Table 1: Table of putative annotations obtained from the MALDI-IMS-qTOF-MSI dataset that could be correlated to molecular features (M_e < 0.06 Da) from the MALDI-TOF MSI analysis highlighted by the ROC analysis in (1a) the embryo; (1b) the endosperm; (1c) the epidermal layer. Rows displaying alternative annotations for a molecular feature of interest from the MALDI-MSI dataset are shown in yellow.

1a	AUC	Measured *m/z* (TOF)	Measured *m/z* (qTOF)	Ion	M_e TOF (Da)	M_e qTOF (ppm)	Measured CCS	Δ CCS %	Name	Molecular formula
Positive ion mode	0.83	160.03	160.0355	$[M+Na]^+$	0.00	8.742	182.2	–	Trigonelline	$C_7H_7NO_2$
	0.89	175.13	175.1177	$[M+H]^+$	−0.01	−7.149	–	–	Arginine	$C_6H_{14}N_4O_2$
	0.95	176.01	176.0090	$[M+K]^+$	0.00	10.431	185.0	–	Trigonelline	$C_7H_7NO_2$
	0.82	181.06	181.0744	$[M+H]^+$	0.02	−13.243	130.6	−0.4	Theobromine	$C_7H_8N_4O_2$
	0.82	181.06	181.0744	$[M+H]^+$	0.02	−13.243	130.6	−3.7	Theophylline	$C_7H_8N_4O_2$
	0.89	195.09	195.0888	$[M+H]^+$	0.00	−5.885	137.5	−1.4	Caffeine	$C_8H_{10}N_4O_2$
	0.88	205.09	205.0985	$[M+H]^+$	0.00	−6.563	138.3	−8.2	Tryptophan	$C_{11}H_{12}N_2O_2$
	0.76	221.07	221.0418	$[M+K]^+$	−0.02	1.792	138.6	–	Chiral alditol (mannitol, galactitol)	$C_6H_{14}O_6$
	0.85	241.18	241.1806	$[M+K]^+$	0.00	−7.028	154.6	–	Spermine	$C_{10}H_{26}N_4$
	0.83	325.14	325.1571	$[M+K]^+$	0.02	−2.036	175.8	–	Androstenedione	$C_{19}H_{26}O_2$
	0.92	330.09	330.0560	$[M+H]^+$	−0.03	11.504	169.0	0.6	AMP cyclic	$C_{10}H_{12}N_5O_6P$
	0.91	355.17	355.1724	$[M+K]^+$	0.00	−15.196	184.5	–	Cafestol	$C_{20}H_{28}O_3$
	0.86	381.14	381.0760	$[M+K]^+$	−0.06	8.841	173.3	−1.2	Disaccharide (sucrose)	$C_{12}H_{22}O_{11}$
	0.96	393.10	393.0551	$[M+K]^+$	−0.04	7.989	181.8	–	Caffeoylquinic acid (CQA) regioisomers	$C_{16}H_{18}O_9$
Negative ion mode	0.82	285.00	285.0402	$[M-H]^-$	0.04	−2.83	157.3	−4.9	Kaempferol	$C_{15}H_{10}O_6$
	0.90	289.00	289.0266	$[M+Cl]^-$	0.03	−1.34	154.8	–	Daidzein	$C_{15}H_{10}O_4$
	0.83	302.99	303.0488	$[M+Cl]^{--}$	0.06	0.77	153.5	–	Formononetin	$C_{16}H_{12}O_4$
	0.90	305.01	305.0187	$[M+Cl]^{--}$	0.02	7.96	166.2	–	Genistein	$C_{15}H_{10}O_5$
	0.80	321.01	321.0117	$[M+Cl]^-$	0.00	13.43	158.8	–	Kaempferol	$C_{15}H_{10}O6$
	0.77	383.05	383.0099	$[M+Cl]^{--}$	−0.03	14.37	185.6	–	Inosine 5 monophosphate	$C_{10}H_{13}N_4O_8P$
	0.83	426.03	426.0228	$[M-H]^-$	−0.01	−4.17	190.6	3.2	ADP	$C_{10}H_{15}N_5O_{10}P_2$
	0.91	475.02	475.0151	$[M+Cl]^{--}$	−0.01	−11.50	200.8	–	Unknown[a]	$C_{18}H_{16}O_{11}S$
	0.80	533.03	533.0877	$[M+Cl]^-$	0.05	−5.97	202.9	–	Dicaffeoyl-quinolactone (di-CQL) regioisomers	$C_{25}H_{22}O_{11}$
	0.82	551.05	551.1004	$[M+Cl]^{--}$	0.04	−9.65	213.9		Dicaffeoylquinic acid (di-CQA) regioisomers	$C_{25}H_{24}O_{12}$
	0.84	563.05	563.0630	$[M-H]^-$	0.02	7.86	229.4	9.5	Deoxythymidine-diphosphate-glucose (dTDP-glucose)	$C_{16}H_{26}N_2O_{16}P_2$
	0.79	571.29	571.2895	$[M-H]^-$	0.00	−2.99	234.2	2.8	LPI 16:0	$C_{25}H_{49}O_{12}P$
	0.81	595.29	595.2895	$[M-H]^-$	0.00	−2.89	235.9	2.2	LPI 18:2	$C_{27}H_{49}O_{12}P$

1b	AUC	Measured *m/z* (TOF)	Measured *m/z* (qTOF)	Ion	M_e TOF (Da)	M_e qTOF (ppm)	Measured CCS	ΔCCS (%)	Name	Molecular formula
Positive ion mode	0.75	758.51	758.5665	$[M+H]^+$	0.06	13.0	286.7	2.2	PC 34:2	$C_{42}H_{80}NO_8P$

1c	AUC	Measured *m/z* (TOF)	Measured *m/z* (qTOF)	Ion	M_e TOF (Da)	M_e qTOF (ppm)	Measured CCS	ΔCCS (%)	Name	Molecular formula
Negative ion mode	0.76	353.10	353.0845	$[M-H]^-$	−0.01	6.2	187.1	9.3	Caffeoylquinic acid (CQA) regioisomers	$C_{16}H_{18}O_9$

[a]Alternative annotations are reported in Supplementary Information, S4.

Table 2: Table of putative annotations obtained from the UHPLC-ESI-IMS-MS/MS dataset that were found to be common with the MALDI-IMS-qTOF-MSI features of interest. When more than one adduct was present for a putative compound, the annotation with the least mass error (ppm) was selected.

Name	Measured *m/z*	Ion	M_{error} (ppm)	Measured CCS	ΔCCS (%)	mSigma	MS/MS score	Molecular formula
Trigonelline	138.05542	$[M+H]^+$	−3.368	122.8	−3.5	1.3	1000.0	$C_7H_7NO_2$
Arginine	175.11889	$[M+H]^+$	−0.335	135.7	−3.0	9.9	1000.0	$C_6H_{14}N_4O_2$
Theophylline	181.07202	$[M+H]^+$	−0.099	135.7	0.9	42.3	583.5	$C_7H_8N_4O_2$
Caffeine	195.0876	$[M+H]^+$	0.267	139.3	−0.6	8.4	677.4	$C_8H_{10}N_4O_2$
Daidzein	255.06256	$[M-H]^-$	−9.955	147.3	–	38.9	–	$C_{15}H_{10}O_4$
Cafestol	339.19349	$[M+Na]^+$	−1.253	194.2	−3.5	4.9	–	$C_{20}H_{28}O_3$
Disaccharide (sucrose)	341.10906	$[M-H]^-$	−3.582	167.8	−0.2	3.3	994.8	$C_{12}H_{22}O_{11}$
Caffeoylquinic acid (CQA) regioisomers	353.08786	$[M-H]^-$	−3.260	170.8	−0.2	2.1	954.5	$C_{16}H_{18}O_9$
Dicaffeoyl-quinolactone (di-CQL) regioisomers	499.12346	$[M+H]^+$	0.056	239.6	–	26.7	1000.0	$C_{25}H_{22}O_{11}$
Dicaffeoylquinic acid (di-CQA) regioisomers	515.11935	$[M-H]^-$	−1.838	213.5	3.6	4.9	960.0	$C_{25}H_{22}O_{11}$
LPI 16:0	571.28942	$[M-H]^-$	−2.853	233.9	0.4	16.9	1000.0	$C_{25}H_{49}O_{12}P$
LPI 18:2	595.28890	$[M-H]^-$	−1.865	235.7	2.1	3.1	1000.0	$C_{27}H_{49}O_{12}P$
PC 34:2	758.56948	$[M+H]^+$	−0.063	286.1	−1.1	38.6	1000.0	$C_{42}H_{80}NO_8P$

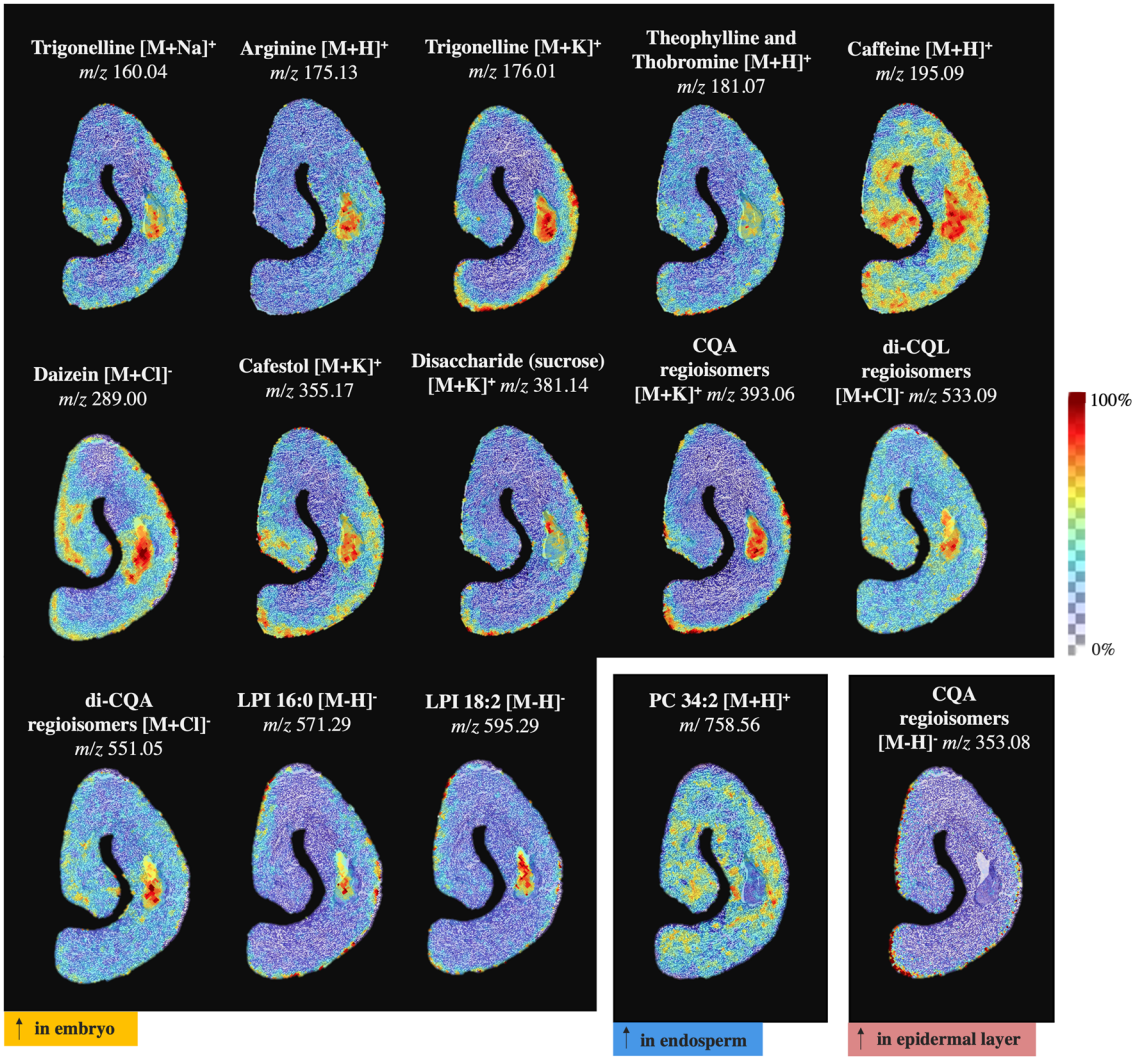

Fig. 3: Representative MALDI-MSI images of 15 compounds whose distribution was enriched with AUC > 0.75 in either the embryo (yellow label), endosperm (blue label) or epidermal layer (red label) of *C. arabica* coffee beans. The distribution of each molecular feature within the tissue section is displayed through an intensity score chart; the intensity of colour is not representative of the abundance of a feature within the spectrum, but rather of the distribution of said feature within the single MALDI-MSI image.

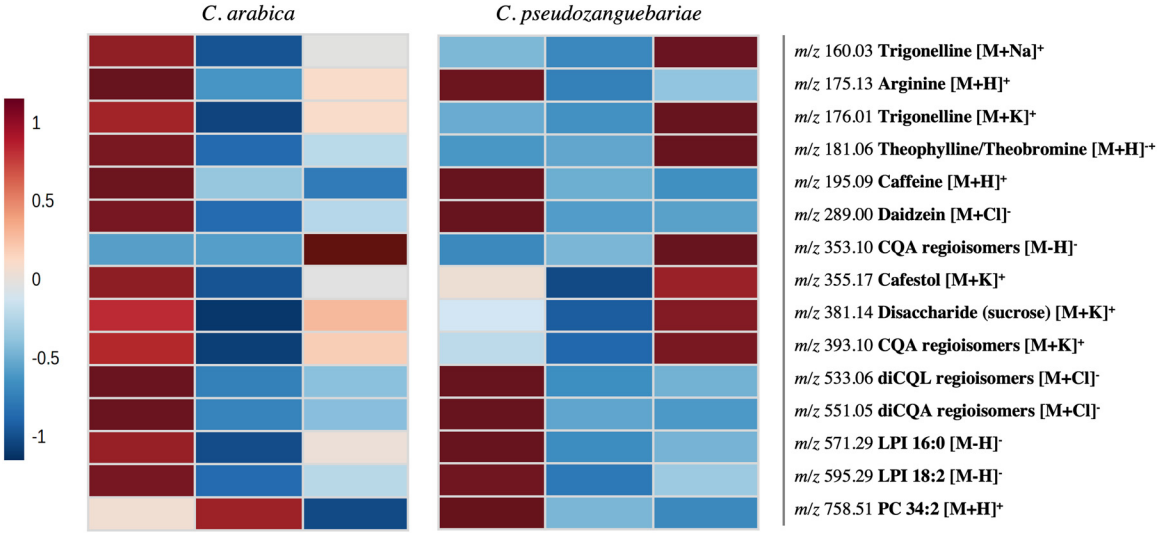

Fig. 4: Heatmaps generated by confronting the mean intensities of 15 molecular features specifically enriched within the embryo, endosperm, or epidermal layer of single *C. arabica* and *C. pseudozanguebariae* bean sections.

C. arabica coffee bean section and a low intensity in other tissues, the same metabolite was instead found in higher abundance in the embryo of the *C. pseudozanguebariae* bean section.

Discussion

In this study, we explored the spatial metabolome of *C. arabica* coffee beans to assess the tissue distribution of molecular species whose localisation mainly contributed to the biochemical composition of the principal tissues of the bean, namely the epidermal layer, the soft endosperm, the hard endosperm, and the embryo. Relying on high spatial and mass resolution techniques whose complementary results can be successfully correlated, including MALDI-TOF MSI, MALDI-IMS-qTOF-MSI, and UHPLC-ESI-IMS-MS/MS, it was possible to assess with high confidence the spatial distribution of 13 compounds within different tissues of a coffee bean section.

In a first instance, coffee bean sections containing the embryo were analysed with a MALDI-TOF MSI approach, correlating the molecular data with histologically stained images in order to obtain annotated regions representing the main bean tissues. Interestingly, the principal component analysis could distinguish clusters derived from the embryo and the epidermal layer, whilst the soft and hard endosperm were grouped in a common cluster, indicating that the two regions present a highly similar metabolome which precludes the possibility to separate them based on their molecular features, despite the anatomical differences and their different cell wall content, as highlighted by Eira *et al.* [4]. These findings could be justified by accounting for the different roles that these tissues hold during seed germination in the genus Coffea, which support the presence of different metabolic profiles. Specifically, large exchanges and diffusion of metabolites are also supposed to occur in the parenchymatous cells of the endosperm, facilitated by the presence of plasmodesmata. This should help the migration of some compounds from the internal endosperm to its epidermal layer during the last stages of seed maturation [6]. These migrations of compounds in coffee seed could explain the similar metabolic content in the hard and soft endosperm region and the different characteristics highlighted in the outer epidermal layer. In line with the findings of the PCA, spatial segmentation indicated the embryo as a distinct region in terms of metabolome; on the other hand, there weren't any clusters that could efficiently separate the soft and hard endosperm, which is another indication that these tissues, although distinguishable through histological staining based on the orientation and coloration of cells, are actually characterised by an analogous metabolomic profile, as already inferred by the PCA. Interestingly, spatial segmentation highlighted the occurrence of concentric metabolic

clusters toward the convex, external portion of the bean, present both in the positive and the negative ion mode datasets, whose distribution did not correlate with any known bean tissues except for a partial overlap with the outer portion of the epidermal layer. A possible reasoning behind these findings could involve the need for the bean to defend against atmospheric events or pathogens during the stage of bean life which precedes germination, which would assume the presence of specific features localised around the bean to serve as defense mechanisms [49].

To further investigate these preliminary results, MALDI-IMS-qTOF-MSI was performed to profile *C. arabica* beans, and 29 low molecular weight molecules could be successfully correlated with features of interest emerged through the ROC analysis performed on the MALDI-TOF MSI. Interestingly, many putatively annotated compounds highlighted by the MALDI-IMS-qTOF-MSI dataset were known to be bioactive. Bioactive compounds carry out physiological functions in the plant, but concurrently act as bioactive compounds often linked to health benefits when consumed, and these molecules include alkaloids, such as caffeine, trigonelline, theophylline, and theobromine, but also chlorogenic acids (CGAs), phytoestrogens such as isoflavones and lignans, carbohydrates, lipids, and more [1, 3, 26]. For instance, among foods, coffee contains the highest abundance of chlorogenic acids, well characterised esters of *trans*-cinnamic and quinic acids whose antioxidant activity is linked to a reduced risk of several chronic diseases, including inflammation, diabetes, cardiovascular, and neurodegenerative diseases [1, 50]. These antioxidant properties are also attributable to caffeine, diterpenoids such as cafestol and kahweol, isoflavones and lignans, quercetin and its derivatives, along with many other contributors to the beneficial effects of coffee consumption and could be found in the putatively annotated MALDI-IMS-qTOF-MSI dataset [26, 51–54]. The presence of several bioactive compounds within the annotated dataset calls for a thorough investigation of their distribution within the bean as their beneficial effect on human health could be improved by obtaining further knowledge regarding their physiological role.

In addition to well-known coffee compounds, Table 1 highlighted analytes which were not yet characterised in *C. arabica* beans. For instance, androstenedione is a steroid hormone which is not typical of coffee but, amongst other androgens and estrogens, has already been found in plants [55]. Whilst these preliminary results are interesting, further studies should be conducted to confirm this putative annotation, which is solely based on mass error. Likewise, dTDP-glucose was found enriched within the embryo, and while several rare nucleotide sugars, including TDP derivatives, have already been found in plants, their roles are not yet known [56]. Furthermore, quinolactones, such as di-caffeoylquinolactone, are only formed from CQAs in coffee beans upon roasting and are virtually absent in *C. arabica* green coffee. Interestingly, another study reports the presence of di-CQL in green *C. arabica*, supporting our findings which indicate the presence of di-CQL in the embryo [57]. Lastly, one feature from Table 1 could not be univocally annotated and was indicated as “unknown”, although one of the possible annotations (see Supplementary Information, S4) is represented by a hydroxycinnamic acid, which would be more plausible in green coffee [12].

Finally, UHPLC-ESI-IMS-MS/MS was performed to add a further degree of confidence to the putative annotations obtained through MALDI-IMS-qTOF-MSI. Consequently, it was possible to visualise the tissue distribution of 15 molecular features which were concurrently present in all datasets. Starting from the most abundant CGAs, which as previously stated are renowned abundant bioactive compounds in coffee beans, it was possible to assess the distribution of two adducts of caffeoylquinic acid regioisomers (CQA $[M-H]^-$, *m/z* 353.10 and CQA $[M+K]^+$, *m/z* 393.02), di-caffeoylquinic acid regioisomers (di-CQA $[M+Cl]^-$, *m/z* 551.05) and a CGA derivate, dicaffeoylquinolactone regioisomers (di-CQL $[M+Cl]^-$, *m/z* 533.02) previously detected in green *C. arabica* [57]. Interestingly, di-caffeoylquinic acid and di-caffeoylquinolactone were ubiquitous, but mostly located towards the embryo of the bean; caffeoylquinic acid adducts were abundant in the epidermal layer, with the adduct $[M+K]^+$ being also enriched in the embryo. The discrepancy in the distribution of $[M+K]^+$ adducts of the same compounds within the coffee section could be dependent on the biodistribution of potassium, which is the most abundant metal ion in coffee seeds, within the bean [30], which could represent a biasing factor. As for the distribution towards the embryo or the epidermal layer of CGAs, this result could be directly linked to the physiological role of these compounds within coffee beans. The roles of acyl-quinic acids in planta are not known with any certainty, but evidence suggests that CQAs may accumulate as a store of cinnamic acid for lignin and phenolic polymers biosynthesis in cotyledonary cell walls after seed germination, thus corroborating their abundance in the

embryo [12, 32]. In other species, such as in tomato, globe artichoke, and chamomile, acyl-quinic acids appear to protect against UV damage, which would support their localisation in the external portion of the bean [12]. In fact, phenolic compounds possess diverse functions in plants, including chemical signalling and defence against invaders; interestingly, a study employing DESI-MSI to assess the changes in CQAs distribution and abundance due to pathogens pointed out that CQA showed increased levels in samples attacked by the insect *H. hampei*, indicating an important role in the response to parasites [32]. Moreover, previous MSI-based studies already hinted at a higher abundance of diCGA in the epidermal layer, further validating our findings [30]. During coffee roasting process, CGAs content decreases due to chemical transformations leading to important volatile aroma compounds and to the formation of their corresponding lactones which contribute to the bitterness of the final beverage [15, 42]. Depending on the intensity of the thermal treatment, and then on the roasting degree, chlorogenic acids may be lost up to 95 % in dark roasted coffee [41]. The observed distribution towards the epidermal layer of CGAs may contribute to expose these compounds to the effects induced by the roasting process.

Whilst a complex mixture of chlorogenic acids was putatively annotated in the MALDI-IMS-qTOF-MSI dataset, only those which are most abundant in coffee were also present in the UHPLC-ESI-IMS-MS/MS dataset. The discrepancy amongst techniques could be attributed to the analytical conditions in which liquid chromatography was carried out, as it was performed with the aim to highlight the most varied classes of compounds to accommodate for the untargeted approach in which this study was conducted. Although most annotations were not confirmed through LC-MS, the molecular maps of putative annotations referring to chlorogenic acids, including those which are only present in traces, are shown in the Supplementary Information, S5.1 and 5.2. Future studies could be directed towards the targeted characterisation of minor chlorogenic acids considering that these compounds have not thoroughly been studied in terms of their role in plant physiology nor bioactive potential [14].

In this study, it was also possible to assess the tissue distribution of the isoflavone daidzein, which was significantly enriched in the embryo, but also present in the epidermal layer, as displayed in Fig. 3. Flavonoids, of which isoflavones are a subclass, have long been known to protect plants from different biotic and abiotic stresses, function as signal molecules and act as antimicrobial agents [36], which, as already hypothesised for CGAs, could be an indication of their presence within the embryo and the external part of the bean. Similar antimicrobial properties have been attributed to alkaloids, and specifically caffeine, further supporting the increased distribution of caffeine in analogous tissues [58]. Moreover, cafestol was found to be enriched in the embryo of *C. arabica* beans, according to previous studies [59]. The presence of kahweol, the other main diterpene in *C. arabica* [60], could be confirmed through both MALDI-IMS-qTOF-MSI and UHPLC-ESI-IMS-MS/MS (Supplementary Information, S6), but it was not possible to correlate its *m/z* with a monoisotopic molecular feature in the MALDI-TOF MSI dataset with and acceptable mass error. The enrichment of trigonelline in the embryo of *C. arabica* could be attributed to its function as reserve of NAD, as hinted in previous research where the relative abundance of trigonelline, nicotinic acid and NAD was investigated in the proximal and distal endosperm and the embryo of *C. arabica* beans [21]; regardless, the increased presence of trigonelline in the epidermal layer of *C. pseudozanguebariae* coffee section suggests it could also exert a protective function, perhaps related to trigonelline's well documented role in defence against UV-B radiation [61, 62]. Additionally, sucrose was found to be enriched within the embryo, but also in the epidermal layer, of the bean. Previous MALDI-MSI studies conducted on peanut seeds found sucrose to be distributed throughout the whole tissue, with a slightly higher ion intensity on the edge of cotyledon, and similar results have been obtained in coffee bean seeds where sucrose was detected across the whole endosperm [32, 34]. In the context of green coffee beans, the increased sucrose concentration within the embryo may be justified by the need for readily available respiratory substrates for the increase in metabolism that accompanies radicle emergence [63]. Lastly, the amino acid arginine was found to be enriched in the embryo region. Previous studies investigated the role of arginine during embryogenesis, concluding that it is one of the main precursors of polyamines, which seem to be directly related to embryogenic competency [64]. Interestingly, amino acids such as arginine and tryptophan, but also sugars, are the most important precursors to the Maillard reaction, which is the main pathway for the production of flavour and aroma during coffee bean roasting [65]. Thus, the characterisation of their presence, abundance, and distribution within the bean is

relevant for the coffee quality in terms of taste and fragrance. The presence of both trigonelline and sucrose in the epidermal layer of the bean makes these aroma/flavour precursors particularly exposed to the effects of the roasting process.

To confirm these preliminary findings obtained by confronting the intensities of metabolites across different *C. arabica* coffee bean tissues, MALDI-TOF MSI was performed on a *C. pseudozanguebariae* coffee bean section. The species *C. pseudozanguebariae* is of particular interest in metabolomic studies involving coffee beans due to its low concentration of caffeine and CGAs, which makes it ideal as a negative control in quantification studies involving these metabolites [48]. It's interesting to underline that the analyte mozambioside, which is one of the few metabolites enriched in the species *C. pseudozanguebariae* compared to *C. arabica* [40], could be successfully annotated through UHPLC-ESI-IMS-MS/MS in this bean. As MALDI-IMS-qTOF-MSI was only performed on *C. arabica*, it was not possible to tentatively annotate the metabolite in this specimen, and thus mozambioside was not included in the results. Partial results regarding the MALDI-TOF MSI and UHPLC-ESI-IMS-MS/MS dataset are reported in Supplementary Information, S7.1 and 7.2. In this study, *C. pseudozanguebariae* was adopted on the premise that the relative abundance of compounds of interest, including caffeine and CGAs, in different compartments of this specimen would better represent their physiological function, considering their natural low abundance. Interestingly, the heatmap referred to *C. pseudozanguebariae* in Fig. 4 displays the presence of caffeine, daidzein, diCQA, di-CQL, LPIs, and PC 34:2 to be enriched within the embryo; on the other hand, trigonelline, theophylline/theobromine, cafestol, disaccharides, and CQA were exclusively present in the epidermal layer. Whilst some of the abundances of said compounds were in agreement with those observed for *C. arabica*, many features were differentially expressed in the embryo, the endosperm, or the epidermal layer depending on the coffee species, thus opening to future considerations upon the biological function of these metabolites.

Through a multi-technique approach in which molecular features found in different datasets could be successfully correlated, this work was able to unravel the spatial metabolome of a *C. arabica* green coffee bean section and, moreover, could assess the distribution of 13 metabolites which were found to be enriched within the embryo, the epidermal layer, or the endosperm of the bean. Notwithstanding some limitations, which regarded the use of a unique section of coffee bean and a limited number of coffee species, and thus further studies should be held to validate these preliminary results, this work still entails relevant implications. Firstly, these results encourage hypotheses regarding the biological relevance of these distributions; for instance, information regarding how an embryo mobilises its internal reserves during the early stages of plant growth could provide insights into the metabolic process of germination, paving the way for studies on how certain crop genotypes germinate better under adverse environmental conditions [36]. Additionally, this work was able to locate which bean tissues were specifically enriched in bioactive compounds, molecules that exert therapeutic effects on human health. These findings highlight the importance of future longitudinal studies where the abundance and distribution of bioactive compounds are assessed in various species and at different stages of seed maturation, which could add further layers of knowledge related to those compounds that contribute to the well acknowledged health benefits linked to coffee consumption.

Conclusions

In this work, we correlated datasets derived from high mass and spatial resolution techniques, namely MALDI-TOF-MSI, MALDI-IMS-qTOF-MSI, and UHPLC-ESI-IMS-MS/MS, to assess which molecules best characterise the main tissues of *C. arabica* green coffee beans, focusing on the yet unexplored embryo region. The integration of these technologies facilitated the detection and annotation of 13 metabolites, of which many are known to be bioactive and aroma precursors. The spatial metabolomic exploration of this coffee specimen allowed us to determine the relative abundance and localisation of these metabolites within the bean and, in turn, support hypotheses regarding the biological activity within this particular matrix as well as the possible exposure to the thermal degradation induced by the roasting process.

Aknowledgments: GB would like to thank Milan Rotary Clubs for partially supporting her with "Borsa di Studio Rotary Dott. Gabriele Corbelli"- Milan Rotary Clubs.
Research funding: The work presented in this manuscript was facilitated by Regione Lombardia: programma degli interventi per la ripresa economica. Sviluppo di nuovi accordi di collaborazione con le università per la ricerca, l'innovazione e il trasferimento tecnologico.

References

[1] B. B. Gökcen, N. Şanlier. *Crit. Rev. Food Sci. Nutr.* **59**, 336 (2019), https://doi.org/10.1080/10408398.2017.1369391.
[2] N. Bhumiratana, K. Adhikari, E. Chambers. *LWT* **44**(10), 2185 (2011), https://doi.org/10.1016/j.lwt.2011.07.001.
[3] F. Bastian, O. S. Hutabarat, A. Dirpan, F. Nainu, H. Harapan, T. B. Emran, J. Simal-Gandara. *Foods* **10**, 2827 (2021), https://doi.org/10.3390/foods10112827.
[4] M. T. S. Eira, E. A. Amaral Da Silva, R. D. De Castro, S. Dussert, C. Walters, J. D. Bewley, H. W. M. Hilhorst. *Coffee Seed Physiol.* **18**, 149 (2006), https://doi.org/10.1590/s1677-04202006000100011.
[5] E. Dentan. The microscopic structure of the coffee bean. In *Coffee*, M. N. Clifford, K. C. Willson (Eds.), Springer, Boston, MA (1985).
[6] R. D. De Castro, P. Marraccini. *J. Plant Physiol.* **18** (2006), https://doi.org/10.1590/S1677-04202006000100013.
[7] E. A. A. Da Silva, P. E. Toorop, A. C. Van Aelst, H. W. M. Hilhorst. *Planta* **220**(2), 251 (2004), https://doi.org/10.1007/s00425-004-1344-0.
[8] R. D. Hall, F. Trevisan, R. C. H. de Vos. *Food Res. Int.* **151**, 110825 (2022), https://doi.org/10.1016/j.foodres.2021.110825.
[9] M. Rosa, F. S. F. R. Alonso-Salces, K. H. Botanical, K HÉberger. *J. Agric. Food Chem*. **57**(10), 4224 (2009), https://doi.org/10.1021/jf8037117.
[10] J. Yan, H. Tong. *Compr. Rev. Food Sci. Food Saf.* **22**(1), 187 (2023), https://doi.org/10.1111/1541-4337.13067.
[11] R. Lang, S. Klade, A. Beusch, A. Dunkel, T. Hofmann. *J. Agric. Food Chem*. **63**(48), 10492 (2015), https://doi.org/10.1021/acs.jafc.5b04847.
[12] M. N. Clifford, I. B. Jaganath, I. A. Ludwig, A. Crozier. *Nat. Prod. Rep.* **34**, 1391 (2017), https://doi.org/10.1039/c7np00030h.
[13] S. Colomban, E. Guercia, L. Navarini. *J. Mass Spectrom.* **55**, e4634 (2020), https://doi.org/10.1002/jms.4634.
[14] S. Colomban, M. De Rosso, R. Flamini, L. Navarini. *J. Mass Spectrom.* **58**, e4970 (2023), https://doi.org/10.1002/jms.4970.
[15] O. Frank, G. Zehentbauer, T. Hofmann. *Eur. Food Res. Technol.* **222**(5–6), 492 (2006), https://doi.org/10.1007/s00217-005-0143-6.
[16] C. Müller, R. Lang, T. Hofmann. *J. Agric. Food Chem*. **54**(26), 10086 (2006), https://doi.org/10.1021/jf062727y.
[17] V. Sinisi, A. Stevaert, F. Berti, C. Forzato, F. Benedetti, L. Navarini, A. Camps, L. Persoons, K. Vermeire. *Planta Med*. **83**(7), 615 (2017), https://doi.org/10.1055/s-0042-119449.
[18] N. Tajik, M. Tajik, I. Mack, P. Enck. *Eur. J. Nutr.* **56**, 2215 (2017), https://doi.org/10.1007/s00394-017-1379-1.
[19] S. F. Nabavi, S. Tejada, W. N. Setzer, O. Gortzi, A. Sureda, N. Braidy, M. Daglia, A. Manayi, S. M. Nabavi. *Curr. Neuropharmacol*. **15**(4), 471 (2016), https://doi.org/10.2174/1570159x14666160325120625.
[20] N. Liang, D. D. Kitts. *Nutrients* **8**, 16 (2015), https://doi.org/10.3390/nu8010016.
[21] M. M. Shimizu, P. Mazzafera. *Plant Biol.* **2**(6), 605 (2000), https://doi.org/10.1055/s-2000-16645.
[22] R. Lang, E. F. Yagar, R. Eggers, T. Hofmann. *J. Agric. Food Chem*. **56**(23), 11114 (2008), https://doi.org/10.1021/jf802838s.
[23] H. W. S. Ashihara, S. Watanabe. *Nat. Prod. Commun*. **9**, 795 (2014).
[24] K. F. Allred, K. M. Yackley, J. Vanamala, C. D. Allred. *J. Nutr.* **139**(10), 1833 (2009), https://doi.org/10.3945/jn.109.108001.
[25] K. Socała, U. Doboszewska, A. Szopa, A. Serefko, M. Włodarczyk, A. Zielińska, E. Poleszak, J. Fichna, P. Wlaź. *Pharmacol. Res.* **172**, 105840 (2021), https://doi.org/10.1016/j.phrs.2021.105840.
[26] S. Angeloni, L. Navarini, G. Khamitova, F. Maggi, G. Sagratini, S. Vittori, G. Caprioli. *Food Chem*. **325**, 126924 (2020), https://doi.org/10.1016/j.foodchem.2020.126924.
[27] Y. Sapozhnikova. *Food Chem*. **150**, 87 (2014), https://doi.org/10.1016/j.foodchem.2013.10.131.
[28] I. Blank, E. C. Pascual, S. Devaud, L. B. Fay, R. H. Stadler, C. Yeretzian, B. A. Goodman. *J. Agric. Food Chem*. **50**(8), 2356 (2002), https://doi.org/10.1021/jf011329m.
[29] A. C. R. Silva, R. Garrett, C. M. Rezende, S. W. Meckelmann. *J. Food Compos. Anal.* **111**, 104587 (2022), https://doi.org/10.1016/j.jfca.2022.104587.
[30] N. Li, J. Dong, C. Dong, Y. Han, H. Liu, F. Du, H. Nie. *J. Am. Soc. Mass Spectrom*. **31**(12), 2503 (2020), https://doi.org/10.1021/jasms.0c00202.
[31] A. R. Buchberger, K. DeLaney, J. Johnson, L. Li. *Anal. Chem.* **90**, 240–265 (2018), https://doi.org/10.1021/acs.analchem.7b04733.
[32] R. Garrett, C. M. Rezende, D. R. Ifa. *LWT* **65**, 711 (2016), https://doi.org/10.1016/j.lwt.2015.08.062.
[33] L. Qin, Y. Zhang, Y. Liu, H. He, M. Han, Y. Li, M. Zeng, X. Wang. *Phytochem. Anal.* **29**, 351 (2018), https://doi.org/10.1002/pca.2759.
[34] X. Wang, Y. Chen, Y. Liu, L. Ouyang, R. Yao, Z. Wang, Y. Kang, L. Yan, D. Huai, H. Jiang, Y. Lei, B. Liao. *Foods* **11**(23), 3888 (2022), https://doi.org/10.3390/foods11233888.
[35] K. Gorzolka, J. Kölling, T. W. Nattkemper, K. Niehaus. *PLoS One* **11**(3), e0150208 (2016), https://doi.org/10.1371/journal.pone.0150208.
[36] S. Gupta, T. Rupasinghe, D. L. Callahan, S. H. A. Natera, P. M. C. Smith, C. B. Hill, U. Roessner, B. A. Boughton. *Front. Plant Sci.* **10**, 1139 (2019), https://doi.org/10.3389/fpls.2019.01139.
[37] I. Flament. *Coffee Flavor Chemistry*, John Wiley & Sons, West Sussex (2002).
[38] C. Campa, S. Doulbeau, S. Dussert, S. Hamon, M. Noirot. *Food Chem*. **91**(4), 633 (2005), https://doi.org/10.1016/j.foodchem.2004.06.032.

[39] C. Campa, S. Doulbeau, S. Dussert, S. Hamon, M. Noirot. *Food Chem*. **93**(1), 135 (2005), https://doi.org/10.1016/j.foodchem.2004.10.015.
[40] F. Anthony, M. N. Clifford, M. Noirot. *Genet. Resour. Crop Evol.* **40**, 61 (1993), https://doi.org/10.1007/BF00052636.
[41] A. Farah, C. M. Donangelo. *Braz. J. Plant Physiol.* **18**, 23 (2006), https://doi.org/10.1590/S1677-04202006000100003.
[42] A. Farah, T. De Paulis, L. C. Trugo, P. R. Martin. *J. Agric. Food Chem*. **53**(5), 1505 (2005), https://doi.org/10.1021/jf048701t.
[43] A. L. Gutiérrez Ortiz, F. Berti, W. Solano Sánchez, L. Navarini, S. Colomban, P. Crisafulli, C. Forzato. *Food Chem*. **286**, 459 (2019), https://doi.org/10.1016/j.foodchem.2019.02.039.
[44] G. Clark. in *Staining Procedures Used by the Biological Stain Commission*, Williams & Wilkins, Baltimore, 4th ed (1981).
[45] H. Hu, R. Yin, H. M. Brown, J. Laskin. *Anal. Chem*. **93**(7), 3477 (2021), https://doi.org/10.1021/acs.analchem.0c04798.
[46] D. P. Overy, D. P. Enot, K. Tailliart, H. Jenkins, D. Parker, M. Beckmann, J. Draper. *Nat. Protoc*. **3**(3), 471 (2008), https://doi.org/10.1038/nprot.2007.512.
[47] V. M. Mendes, M. Coelho, A. R. Tomé, R. A. Cunha, B. Manadas. *Molecules* **24**(16), 2863 (2019), https://doi.org/10.3390/molecules24162863.
[48] C. Bertrand, M. Noirot, S. Doulbeau, A. De Kochko, S. Hamon, C. Campa. *Plant Sci.* **165**(6), 1355 (2003), https://doi.org/10.1016/j.plantsci.2003.07.002.
[49] M. D. C. Silva, L. Guerra-Guimarães, I. Diniz, A. Loureiro, H. Azinheira, A. P. Pereira, S. Tavares, D. Batista, V. Várzea. *Agronomy* **12**(2), 326 (2022), https://doi.org/10.3390/agronomy12020326.
[50] R. Upadhyay, L. J. Mohan Rao. *Crit. Rev. Food Sci. Nutr.* **53**(9), 968 (2013), https://doi.org/10.1080/10408398.2011.576319.
[51] B. E. Ősz, G. Jîtcă, R. E. Ștefănescu, A. Pușcaș, A. Tero-Vescan, C. E. Vari. *Int. J. Mol. Sci.* **23**, 13074 (2022), https://doi.org/10.3390/ijms232113074.
[52] M. Imran, M. S. Arshad, M. S. Butt, J. H. Kwon, M. U. Arshad, M. T. Sultan. *Lipids Health Dis.* **16**, 1 (2017), https://doi.org/10.1186/s12944-017-0449-y.
[53] C. Cárdenas, A. R. Quesada, M. A. Medina. *PLoS One* **6**(8), e23407 (2011), https://doi.org/10.1371/journal.pone.0023407.
[54] A. M. Mustafa, D. Abouelenein, S. Angeloni, F. Maggi, L. Navarini, G. Sagratini, A. Santanatoglia, E. Torregiani, S. Vittori, G. Caprioli. *Foods* **11**(19), 3033 (2022), https://doi.org/10.3390/foods11193033.
[55] A. Janeczko. *Plants* **10**, 2783 (2021), https://doi.org/10.3390/plants10122783.
[56] S. Mikkola. *Molecules* **25**, 5755 (2020), https://doi.org/10.3390/molecules25235755.
[57] A. Farah, T. De Paulis, D. P. Moreira, L. C. Trugo, P. R. Martin. *J. Agric. Food Chem*. **54**(2), 374 (2006), https://doi.org/10.1021/jf0518305.
[58] Y. S. Kim, Y. E. Choi, H. Sano. *Plant Signal. Behav.* **5**, 489 (2010), https://doi.org/10.4161/psb.11087.
[59] V. Sridevi, P. Giridhar, G. A. Ravishankar. *Curr. Sci*. **99**, 1101 (2010).
[60] F. J. M. Novaes, F. C. Bayan, F. R. D. A. Neto, C. M. Rezende. *Coffee Sci.* **14**, 265 (2019), https://doi.org/10.25186/cs.v14i2.1571.
[61] D. Willmon, A. R. Devireddy, M. Inupakutika, N. Puppala, Y. Cho. *Am. J. Plant Sci*. **8**(5), 998 (2017), https://doi.org/10.4236/ajps.2017.85066.
[62] S. Naikoo, S. A. Tasduq. *J. Photochem. Photobiol. B* **202**, 111720 (2020), https://doi.org/10.1016/j.jphotobiol.2019.111720.
[63] K. J. Steadman, H. W. Pritchard, P. M. Deyf. *Tissue-Specific Soluble Sugars in Seeds as Indicators of Storage Category*, Vol. **77** (1996). https://academic.oup.com/aob/article-abstract/77/6/667/2389843.
[64] P. M. C. Bartos, H. T. Gomes, L. I. V. do Amaral, J. B. Teixeira, J. E. Scherwinski-Pereira. *3 Biotech* **8**(4), 209 (2018), https://doi.org/10.1007/s13205-018-1238-7.
[65] Z. Li, B. Zhou, T. Zheng, C. Zhao, X. Shen, X. Wang, M. Qiu, J. Fan. *Foods* **12**(7), 1432 (2023), https://doi.org/10.3390/foods12071432.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/pac-2024-0210).