Accepted Manuscript

Title: Combined multivariate data analysis of high-performance thin-layer chromatography fingerprints and Direct Analysis in Real Time mass spectra for profiling of natural products like propolis

Authors: Gertrud E. Morlock Petar Ristivojevic Elena S. Chernetsova

PII: S0021-9673(13)01934-1
DOI: http://dx.doi.org/doi:10.1016/j.chroma.2013.12.053
Reference: CHROMA 354999

To appear in: *Journal of Chromatography A*

Received date: 12-11-2013
Revised date: 16-12-2013
Accepted date: 17-12-2013

Please cite this article as: G.E. Morlock, P. Ristivojevic, E.S. Chernetsova, Combined multivariate data analysis of high-performance thin-layer chromatography fingerprints and Direct Analysis in Real Time mass spectra for profiling of natural products like propolis, *Journal of Chromatography A* (2013), http://dx.doi.org/10.1016/j.chroma.2013.12.053

Combined multivariate data analysis of high-performance thin-layer chromatography fingerprints and Direct Analysis in Real Time mass spectra for profiling of natural products like propolis

Gertrud E. Morlock[a,*], *Petar Ristivojevic*[a,1], *and Elena S. Chernetsova*[a,2]

[a]**Chair of Food Science, Institute of Nutritional Science, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, and Institute of Food Chemistry, University of Hohenheim, Garbenstrasse 28, 70599 Stuttgart, Germany**

*Corresponding author at: Justus Liebig University Giessen, Institute of Nutritional Science, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. Tel.: +49 641 9939141; fax: +49 6419939149.

E-mail addresses: Gertrud.Morlock@ernaehrung.uni-giessen.de (G.E. Morlock), ristivojevic@chem.bg.ac.rs (P. Ristivojevic), chern_es@mail.ru (E.S. Chernetsova)

[1]On leave from Innovation Center of the Faculty of Chemistry, University of Belgrade, P. O. Box 51, 11158 Belgrade, Serbia

[2]On leave from Research and Innovation Department, People's Friendship University of Russia, Miklukho-Maklaya st. 6, 117198 Moscow, Russia

Abstract

Sophisticated statistical tools are required to extract the full analytical power from high-performance thin-layer chromatography (HPTLC). Especially, the combination of HPTLC fingerprints (image) with chemometrics is rarely used so far. Also, the newly developed, instantaneous Direct Analysis in Real Time mass spectrometry (DART-MS) method is perspective for sample characterization and differentiation by multivariate data analysis. This is a first novel study on the differentiation of natural products using a combination of fast fingerprint techniques, like HPTLC and DART-MS, for multivariate data analysis. The results obtained by the chemometric evaluation of HPTLC and DART-MS data provided complementary information. The complexity, expense, and analysis time were significantly reduced due to the use of statistical tools for evaluation of fingerprints. The approach allowed categorizing 91 propolis samples from Germany and other locations based on their phenolic compound profile. A high level of confidence was obtained when combining orthogonal approaches (HPTLC and DART-MS) for ultrafast sample characterization. HPTLC with selective post-chromatographic derivatization provided information on polarity, functional groups and spectral properties of marker compounds, while information on possible elemental formulae of principal components (phenolic markers) was obtained by DART-MS.

Keywords

Planar chromatography; High-performance thin-layer chromatography; DART-MS; Fingerprint; Pattern recognition; Propolis

1. Introduction

High-performance thin-layer chromatography (HPTLC) is a commonly used technique for screening of herbal liquid extracts [1-3]. Compared to high-performance liquid chromatography (HPLC), and to some degree, to gas chromatography (GC) [4], the interdisciplinary approach using chemometric techniques for data evaluation is rarely employed for TLC or HPTLC [5-9]. Another highly promising technique for chemometric, fingerprint-based sample characterization is Direct Analysis in Real Time mass spectrometry (DART-MS) [10, 11]. Due to its novelty, only a few papers have been published on chemometric studies, proving the method's attractiveness in this field: metabolomics studies for clinical diagnostics [12, 13], origin assignment of beer [14], characterization of olive oils [15], authentication of animal fats [16] and *Umbelliferae* medicinal herbs [17]. The concept of DART-MS analysis is different from that of HPTLC. It is impossible to analyze several samples at the same time, but due to the rapidness of DART-MS (only a few seconds per analysis of one sample) any changes of the analytical conditions are minimized from one sample to another with the exception of matrix interferences. In the majority of cases, HPLC or GC analyses of one sample lasts usually 15-30 min, therefore changes in the analytical conditions may cause a drift in the results [18, 19]; this can successfully be avoided using HPTLC or DART-MS analysis. In HPTLC, up to 72 samples can be analyzed simultaneously under identical conditions, and thus the chromatographic separation of samples takes less than a minute per sample, often only 20 seconds [20-23], which is an important prerequisite for maximizing the outcome of further multivariate data analyses.

The concurrent use of HPTLC and DART-MS might provide an increased confidence level of multivariate data analysis because these independent methods deliver complementary information on the elucidated principal components. In HPTLC, one can get the information

about the analytes due to their hR_F values (evaluation of analytes' polarity), spectral properties (absorbance, fluorescence) and selective derivatization using different reagents (getting knowledge on functional groups); while DART-MS provides the *m/z* values, which can be further transferred into exact molecular weights and suggested elemental formulae in case of high-resolution mass spectrometry. Moreover, after chemometric evaluation of HPTLC chromatograms exposing principal component zones, it is possible to couple HPTLC *online* or *offline* with different mass spectrometric techniques for identification of the components of interest [24].

Propolis (bees glue) is produced by bees while adding their saliva secreted to the resinous plant exudates. Subsequently the partially digested plant material is mixed with wax. In a recent study [25], the HPTLC fingerprints of more than 100 German propolis samples from different locations were visually compared with each other and with respective propolis fingerprints of other countries. Based on this study, mainly two types of German propolis were elucidated, which had a characteristic blue or orange pattern of HPTLC zones. This means that already the visual comparison of HPTLC fingerprints, or analogously shown for DART-MS spectral fingerprints [26], could be used for differentiation without any statistical evaluation. However, some of the German propolis samples were classified as mixed type, because they had zones characteristic for both sample types, and their unambiguous assignment was challenging for analysts and additionally time-consuming. Chemometric techniques were explored in this study for their potential to support this challenging decision process on such mixed type assignments. Although at the very end of our study, a Romanian propolis study by TLC and hierarchical fuzzy clustering was published [27], to our knowledge, the current study is the first one with a novel concept: Until now, there are no publications on the combined use of HPTLC and DIP-it DART-MS for fast pattern recognition and categorization of any samples by chemometric techniques.

2. Material and methods

2.1 Reagents and chemicals, sample preparation and HPTLC

The newly developed HPTLC method for determination of phenolic compounds in propolis extracts (extracted with ethyl acetate) was employed [25]. Briefly, the propolis extracts were obtained from the Apicultural State Institute (Stuttgart, Germany) and were applied on HPTLC plates silica gel 60, 20 x 10 cm, as 8 mm bands. Additionally, a standard mixture was applied described elsewhere [25]. The chromatographic separation was performed in the Twin Trough Chamber 20 x 10 cm (CAMAG, Muttenz, Switzerland) with a mixture of *n*-hexane, ethyl acetate and acetic acid (5:3:1, v/v/v) up to a migration distance of 65 mm (from the lower plate edge). During separation the plate was conditioned with hydrochloric acid (37 %) contained in the second trough of that chamber. Then, the plate was dried under a stream of warm air for 3 min. For derivatization, the plate was immersed in the natural product reagent using the Chromatogram Immersion Device (CAMAG), dried again, then dipped in polyethylene glycol and documented under UV 366 nm using the TLC Visualizer (CAMAG). The plate image was captured by a Baumer Optronic DXA252 digital camera with a 12-bit per channel color depth charge-coupled device (CCD), a 100-µm spatial resolution and an image size of 1922 x 952 pixels. The following capture settings were used: 1500 ms exposure time and gain of 1. The spot amplification tool was used to optimize the visual zone intensity on the images.

104

105 **2.2 Direct Analysis in Real Time mass spectrometry (DART-MS)**

106 For registration of DART mass spectra of propolis extracts, the *DART-SVPA* ion source
107 (IonSense, Saugus, MA, USA) was used. The *DART-SVPA* ion source was equipped with a
108 motorized rail and the 12 DIP-it® sampler, allowing the propolis extracts on 12 glass DIP-it tips
109 to be supplied sequentially into the ionization region with controlled speed. An additionally
110 installed external flowmeter (Analyt-MTC) recorded the helium flow rate through the DART ion
111 source, which was found out to be *ca*. 3.2 - 3.4 L min^{-1} and varied only slightly with the
112 temperature. The DART ion source was operated at 150 and 300 ^{o}C using the DART Control
113 software (IonSense). The movement of a motorized rail was manually controlled with this
114 software using the "move-to-the-left" and "move-to-the-right" buttons in the software program
115 and selecting the speed of movement in the range of 0.2–10 mm s^{-1}. The needle voltage was 4
116 kV, and the voltages at electrodes 1 and 2 were -100 and -250 V, respectively. For operation,
117 helium gas (purity of 99.999 %) was employed, whereas nitrogen gas (99.999 %) was used in the
118 standby mode. The ion source was coupled to the G1956B MSD single quadrupole mass
119 spectrometer (Agilent, Waldbronn, Germany) via the Vapur vacuum interface (IonSense). The
120 evacuation was performed using a Diaphragm Vacuum Pump MZ 2 (Vacuubrand, Wertheim,
121 Germany). The mass spectrometer was operated in the negative ion mode. For data acquisition
122 and processing, the LC/MSD Chemstation B.02.01-SR1(260) software (Agilent) was used. The
123 total ion current (TIC) was registered in the range of *m/z* 70 - 700.

124 **2.3 Multivariate statistical analysis**

125 Principal component analysis (PCA) and hierarchal cluster analysis (HCA) were performed
126 using the PLS Toolbox (demo version 5.7; Eigenvector Research, Wenatchee, WA, USA) for

MATLAB (version 7.8.0 R2011a); MathWorks, Natick, MA, USA). Linear discriminant analysis (LDA) was performed using SPSS 20.0 Statistics software (SPSS Inc., 2012). The data were pre-processed using the auto-scale function in MATLAB.

For the multivariate analysis of HPTLC data, the HPTLC plate images were exported from the winCATS software (CAMAG) to MATLAB. The images were converted into double precision (im2double). After that, the RGB-scale images were converted into grayscale images by eliminating the hue and saturns and retaining the luminance. By means of these operations, each HPTLC image was converted into the data matrix. Obtained data matrices were transferred into Microsoft Office Excel 2007 in order to generate HPTLC profiles of all samples and standard mixtures by calculating R_F values of the target zones and accompanying intensities for each sample's image. Multivariate processing (PCA, HCA and LDA) of obtained HPTLC profiles were performed by NCSS Statistical Software and MATLAB. Numerical values of variables were obtained by calculating the mean values of each zone, separately for all samples. In that case the data matrix was composed of averaged intensities on targeted R_F values as independent variables for each object considered. Chromatograms obtained in this study, have little or no hR_F shifts differences in the hR_F values and chromatographic fingerprints according to visual assignment, which was further confirmed with good chemometrics models [28].

In case of DART-MS, the mass spectral data were directly exported from the ChemStation MS software (Agilent) into Microsoft Office Excel 2007 as CSV files. Prior to carrying out the multivariate statistical analysis, potential marker signals with highest relative abundance (*e.g.,* the *m/z* signal intensity values (%) of the supposed protonated molecules) were chosen.

The unsupervised methods (PCA and HCA) were used for classification of propolis according to orange and blue types. PCA transforms the original, measured variables into new uncorrelated

variables called principal components. Cluster analysis (CA) was used to identify similarity groups between propolis samples. LDA as supervised method was applied to distinguish in groups a collection of propolis samples, having a set of cases whose group membership is already known. It was used for distinguishing among the groups and to develop a procedure for predicting group membership for new cases, finding an optimal decision rule for the classification of the groups.

3. Results and discussion

In the current study, the chemometric techniques were used as an alternative way for classification of HPTLC fingerprints of propolis extracts, and also for classification of propolis extracts based on their DART mass spectra. This double-sided approach for sample characterization increased the level of confidence because different sample characteristics (polarity, functional groups, spectral characteristics, ionization property, molecular weight, *etc.*) were employed for the characterization. Additionally, HPTLC and DART-MS provided complementary information on the nature of the compounds of interest.

3.1 Pattern recognition based on HPTLC fingerprints

In herbal analysis, pattern recognition is usually based on the combination of hyphenated techniques with chemometric tools [29, 30], whereas the multivariate analysis of HPTLC plate images is rarely used [9, 27]. In the majority of publications the HPTLC fingerprints were only used for visual comparison, not supported by chemometric tools. Combining HPTLC with multivariate data analysis is a promising field of research in herbal analysis, in which the knowledge on important sample components can be gained and deepened.

3.1.1 Principal component analysis (PCA)

PCA is the most often used technique in multivariate analysis, allowing eased visualization of all information contained in a data set. Using PCA as a projection method, the multidimensional data set is transformed into 2D or 3D coordinates. PCA classifies samples according to similarity, determines objects showing different properties from others (outliers), and defines important variables for classification [31]. In the current study, PCA was performed on data sets of fingerprints obtained from the HPTLC plate images. Three principal components (PCs) described 84.16 % of the total data variability. PC1 described 44.96 % of the variability, while PC2 and PC3 described 30.50 % and 8.70 %, respectively. Five samples were outside the Hotelling T^2 95% probability ellipse and therefore removed from the dataset as outliers. Thus 59 samples were considered out of 64 samples. There were two general clusters on the 2D PC-score (Fig. 1), which corresponded to the two groups of the orange and blue type samples. Kaempferol, naringenin and caffeic acid had a positive correlation with PC1. Both types of propolis contained chrysin and quercetin, but their concentration in the orange type samples was significantly higher. According to Fig. 1 PC2 showed a negative correlation with galangin, naringenin, caffeic acid and the unknown compound M4 (assigned as pinobanksin based on our recent MS studies [32, 33]); while apigenin and ellagic acid showed a positive correlation. Also, the HPTLC zone of chrysin was more intensively colored in orange than in blue type samples (Fig. 2), which was confirmed with multivariate data analysis. Further, galangin, kaempferol and naringenin were characteristic for the orange type propolis. To conclude, most of the used standard compounds were characteristic for orange type samples and positioned on the negative PC2-score, like the HPTLC zones of galangin and caffeic acid .

3.1.2 Hierarchical cluster analysis (HCA)

HCA divides all samples into groups (clusters) according to similarity and finds the similarity among objects in a multidimensional space, forming clusters between the nearest objects. There

196 are several ways to determine the distance among the samples in multivariate space. Cluster
197 analysis examines the inter-point distance, and represents this distance into a two-dimensional
198 dendrogram [34, 35]. Thus, the distance between two points in the dendrogram, determines the
199 similarity or dissimilarity among these objects. Using cluster analysis, the samples were grouped
200 according to similarity in a multidimensional system. The best results for HCA were obtained
201 using the Ward method as a tool for calculating the cluster distances and applying the Euclidean
202 distance for measuring the distance between samples. HCA was performed using the HPTLC
203 images (chromatograms) of 59 samples, of which 27 were of the blue type, and 32 of the orange
204 type. Three clusters at a similarity level of 20 % were obtained for the propolis samples (Fig.
205 3A). The right cluster contained 15 blue type propolis samples; the left cluster consisted of 24
206 orange and 1 blue type samples; and the middle cluster comprised 11 blue and 8 orange type
207 samples. The left cluster contained samples with a high content of ellagic acid, kaempferol and
208 quercetin, and showed a different pattern. Some of these samples are positioned on the PCs score
209 outside or near to the Hoteling ellipse or between orange and blue groups.
210 The cluster analysis of the phenolic markers showed three clusters at a similarity level of 15 %
211 (Fig. 3B). Apigenin, quercetin and ellagic acid formed one cluster. Apigenin and quercetin were
212 characteristic for the blue propolis type, while the content of quercetin, which was positioned
213 between the orange and blue cluster, was higher in the orange propolis type. Both other clusters
214 contained phenolic markers, which were characteristic for the orange propolis type. These results
215 were in good agreement with the results obtained in PCA.

216 **3.1.3 Linear discriminant analysis (LDA)**

217 LDA is the commonly used supervised method, determines the function, minimizing the ratio of
218 within-class variance and maximizing the ratio of between-class variance [14, 31, 38].

Recognition and prediction abilities representing the percentage of correctly classified investigated samples during model training and cross-validation were performed. According to the Fisher criterion, the number of discriminant functions found is equal to the number of classes minus one, if the number of variables is larger than the number of classes [39]. The calculated standardized canonical coefficients identify the variables that are important for distinguishing the groups and developed a procedure for predicting a group membership for new cases, finding an optimal decision rule for the classification of a group. LDA was performed on a training set consisting of 59 propolis samples and on a test set contained additional 27 samples of other (unclear) propolis types: propolis samples of the mixed orange-blue type and foreign propolis samples with a different pattern of colored zones on the HPTLC plate. Variables with high weighting in principal components were selected for the LDA model. According to the standardized canonical coefficients (Table S-1), variables with the highest discriminating power are: caffeic acid, naringenin, apigenin and quercetin. The data obtained by all three chemometric methods were in good agreement, and the LDA results confirmed the good differentiation for both propolis types within the training set (Fig. 4 A). The overall correct classification rate was 96.6 % using the original and 91.5 % using the cross validation methods (Table 1). The blue type of propolis was classified with slightly better accuracy (96.3 %) than the orange type (87.5 %). Linear discriminant scores of the studied test samples were calculated according to the equations of linear discriminant scores (Equations S-1 and S-2). Mutual projection of the test samples scores was further presented along with the scores of the training set samples (Fig. 4 B). It can be concluded that most of the foreign samples are grouped separately of both, orange and blue type propolis owed to the different flora in the foreign countries. Also, most of the unclear mixed test

241 samples are statistically assigned to the blue propolis type. The marker compounds were the
242 same, but their content decided on the assignment to the blue or orange type sample.

243 **3.2 Pattern recognition based on DART-MS fingerprints**

244 The optimal temperature used for DART ionization varied mainly between 150 and 300 °C in
245 literature, depending on the samples analyzed and analytes of interest [10]. Therefore, two
246 different DART temperatures, *i.e.* 150 and 300 °C, were preliminarily investigated for their
247 impact on multivariate data analysis using a limited set of propolis extracts (8 blue and 8 orange
248 type samples previously classified by HPTLC). The intention was to determine favorable
249 conditions for further extended classification studies. The classification based on statistical
250 approaches and mass signal compositions of phenolic compounds failed at the temperature of
251 150 °C due to the low signal abundance of the marker compounds and the relatively high
252 background. The relative intensity of characteristic ions of the two propolis types was low and
253 insufficiently for multivariate data analysis. However, using the increased temperature of 300 °C,
254 the samples were successfully classified as either orange or blue types (Fig. 5) using the three
255 statistical approaches (PCA, HCA and LDA), and hence, the temperature of 300 °C was used for
256 the subsequent statistically supported pattern recognition.

257 **3.2.1 Principal component analysis (PCA)**

258 PCA was performed on the initial data (relative abundance, *m/z* ratio) using a covariance matrix
259 with auto scaling. This resulted in a three-component model which explained 68.96 % of total
260 data variability. The principal component PC1 described 38.50 % of the total variation, whereas
261 PC2 and PC3 contributed to 21.14 and 9.32 % of the total variability, respectively, which is
262 illustrated in the 2D-score value plot and the projection of loading vectors for the first two
263 principal components (Fig. 6). The respective *m*/*z* values of marker compounds as variables

have been classified according to their contributions to PC 1 and PC2, and assumptions of respective component structures have been made, where possible (Table 2). In particular, the samples of the orange propolis type, localized in the left side of the PC score plot, contained mainly components with mass signals at *m/z* 517.1, 541.1, and 583.1. These signals could be related to glycosides or phenolic dimers. The blue type propolis samples, positioned on the right side of the PC score, implied mainly phenolic compounds with mass signals at *m/z* 163.1, 242.1, 253.1, 327.1, 343.1, and 417.1 (Fig. 6). The mass signal at *m/z* 327.1 could be the deprotonated molecule of pinobanksin-5-methylether-3-O-acetate or pinobanksin-3-O-propionate [36]. The mass signal at *m/z* 285.1 could be the deprotonated molecule of kaempherol, pinobanksin-5-methylether [36] or luteolin [37]. According to further ESI-MS literature [40-42] mass signals at *m/z* 163.1 and 255.1 indicate the deprotonated molecule of coumaric acid and pinocembrin or liquiritigenin, respectively. Further mass signals at *m/z* 269.1 could be the deprotonated molecule of galangin or apigenin or pinostrobin or benzyl caffeate and at *m/z* 271.1 of naringenin or pinobanksin. The mass signal at *m/z* 253.1 corresponded to the deprononated molecule of chrysin. All these DART mass signals were present in the orange propolis type and positioned on the upper left side score. As blue type assigned propolis samples in the upper middle half of score plot mainly contain flavonoids with signals at *m/z* 269.1, and 285.1, which indicated the deprotonated molecules of galangin or apigenin, and kaempferol. Both types of propolis contained mass signals at *m/z* 269.1, 332.1, and 449.1, located between orange and blue type clusters. In a recent high resolution DART mass spectrometry study [32], the presence of coumaric acid was confirmed in German propolis extracts, but for the *m/z* values of 253.1 and 255.1 not only chrysin and pinocembrin were assigned, but also methoxyflavanone and

liquiritigenin isomers. For *m/z* 271.1, not only naringenin but also pinobanksin was assigned. Besides, for *m/z* 269.1 not galangin or apigenin was assigned, but pinostrobin or benzyl caffeate.

3.2.2 Hierarchical cluster analysis (HCA)

The recognition of the two propolis sorts by using cluster analysis was satisfactory. There were 59 samples of propolis in total, 33 of them were orange type and 26 of them blue type. One cluster contained mostly orange type propolis samples, namely 32 orange and 4 blue type samples, while the second cluster contained 1 orange and 22 blue type samples of propolis (Fig. 7 A). In the HCA dendrogram, the 59 propolis samples were clustered into two groups at a similarity level of 30 % according to their phenolic markers. The cluster of the orange type propolis samples was divided into two smaller clusters at a similarity level of 16 %. The blue type sample cluster was divided into two clusters with 15 % of similarity level. Some of the blue type propolis samples are present in the orange type propolis cluster and show similarity with them, because they contain small amounts of phenolic compounds, which were also characteristic for the orange type propolis samples.

According to the similarity level of 15 % in the dendrogram of the variables, there were three main clusters (Fig. 7 B). The variables of *m*/*z* 517.2, 541.1 and 583.1 (top cluster in Fig. 7 B) showed similarity and appeared in the orange type samples. The second cluster group (clusters in the middle of Fig. 7 B) contained deprotonated molecules, characteristic for the orange type of propolis. The variables of *m*/*z* 163.1, 242.1, 253.1, 285.1, 327.1, 343.1 and 417.1 formed the third cluster (bottom cluster in Fig. 7 B) and determined the blue type propolis samples in good agreement with the results of the PCA scoring. These results confirmed the importance of the phenolic compounds composition for the propolis classification system.

3.2.3 Linear discriminant analysis (LDA)

Similarly to LDA for HPTLC images, LDA for DART mass spectra was performed on a training set consisting of 59 propolis samples and on a test set, which contained additional 27 samples of other propolis types (11 foreign and 16 mixed orange-blue types of propolis). Recognition and prediction abilities representing the percentage of correctly classified samples of propolis during model training and cross-validation were calculated. Linear discriminant scores of test samples analyzed by DART were calculated according to the linear discriminant equations' scores (Equations S-3 and S-4). Variables with a high weight for the first three PCs (253.1, 271.1, 343.1, 417.1, 517.2, 541.1, and 583.1) were chosen for the first step in LDA. According to the standardized canonical coefficients (Table S-2), the variables at *m*/*z* 343.1, 517.2, 583.1 and 327.1 were discriminating blue and orange propolis types. The successful separation among blue and orange type propolis samples is clearly demonstrated for the training set (Fig. 8A), and the mutual projection of the test sample scores was further presented along with the scores of the training set samples. Correct classification (96.6%) was obtained by using the original and 94.9 % using the cross validation method. The orange type was classified with slightly better accuracy (97.0 %) than the blue type (92.3 %, Table 3). Most of the mixed orange-blue type samples in the test set could clearly be assigned to the orange and blue propolis cluster according to the training set. Also, the majority of the foreign propolis samples in the test set could be classified as orange propolis type and only three were assigned to the blue propolis type (Fig. 8 B). Similar to the HPTLC fingerprints, some of the foreign samples could not be assigned as these showed a different pattern owed to the different flora in the respective country.

3.3 Benefit of combined pattern recognition by DART-MS and HPTLC

To conclude, the combination of HPTLC fingerprints with DART mass signals led to the successful separation of orange and blue propolis samples and turned out to be a fast, efficient

and low-cost method for profiling and characterization of natural products such as propolis. The two fingerprint techniques were considered as complementary to each other with regard to the information they provide and its combination improved the reliability of the assignments obtained. According to the PCA model obtained by the HPTLC fingerprint, galangin, kaempferol and naringenin were characteristic for the orange type propolis, while gallic acid and apigenin are characteristics for blue propolis. Also, chrysin was more intensively colored in orange than in blue type samples. The PCA model of the HPTLC fingerprint described more of the total variability with the first three components than the DART fingerprint model. HCA showed a good separation between both propolis types, and all variables present in the orange type formed one cluster, characteristic for the orange type. Therein, galangin, kaempferol and chrysin were positioned close to each other in the dendogram and highly characterized the orange type. According to the standardized canonical coefficients in the LDA model, variables with the highest discriminating power were: caffeic acid, naringenin, and quercetin. Pattern recognition based on the DART fingerprint determined mass signals characteristics for the fingerprint of propolis. The PCA model showed that samples of the orange propolis type mainly contained components with mass signals at *m/z* 517.2, 541.1, and 583.1. The opposite of that, other mass signals such as 163.1, 242.1, 253.1, 327.1 and 343.1, were characteristic for the blue propolis type. Similarly to PCA, the cluster analysis showed a separated cluster with *m*/*z* 517.1, 541.1 and 583.1 for the orange type samples. According to the standardized canonical coefficients in the LDA model, the variables at *m*/*z* 343.1, 517.2, 583.1 and 327.1 were discriminating blue and orange propolis types.

4. Conclusions

A novel combination of analytical methods, namely HPTLC and DART-MS, was introduced for multivariate data analysis (PCA, HCA and LDA). The combination significantly reduced the

relevant effort for pattern recognition and categorization of samples which turned out to be time-consuming especially for unclear sample assignments. Exemplarily, 64 propolis samples from different locations were analyzed, and then the trained system was applied to 27 unknown samples. As results of this statistically supported classification of 91 samples, a profound decision on the assignment of critical, mixed or foreign sample types was obtained. HPTLC and DART-MS proved to be beneficial methods in combination with chemometric techniques as time-dependent variances were mitigated by these fast fingerprint techniques. This novel approach shows great potential for further improvements and integration of other statistical techniques and image evaluation tools. The ease of analysis and the analytical capacity offered by HPTLC and DART encourage their adoption as a common powerful analytical fingerprint technique, especially in combination with chemometric tools for pattern recognition. Their combination could also be transferred to issues of biological and geographical origin of food products. For future studies, the coupling of DART with high-resolution mass analyzers is highly promising for the DART-MS-based fingerprinting, as it will allow not only finding the characteristic *m*/*z* values, but also suggesting the elemental formulae for them for identifying the marker compounds, typical for certain types of samples, with the higher degrees of confidence. This will strengthen the capabilities of the presented combined HPTLC/DART-MS approach, described in the current study.

Acknowledgment

The authors thank Nadine Kunz and Dr. Annette Schroeder, both Apicultural State Institute, Stuttgart, Germany, for collecting the propolis samples and preparation of the propolis extracts, to Irina Scholl for the HPTLC fingerprints of the propolis extracts performed during her diploma

thesis supervised by G. Morlock, and to WALA Heilmittel GmbH for initiating the preliminary project. Thanks to IonSense, Saugus, USA, CAMAG, Muttenz, Switzerland, and Merck Millipore, Darmstadt, Germany, for equipment and plates. This work was financially supported by the program Erasmus Mundus Action 2 "IAMONET-RU".

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.xxx. Tables S-1 and S-2 show standardized canonical coefficients of variables chosen for LDA analysis of HPTLC fingerprints as well as of DART-MS fingerprints. Equations S-1, S-2, S-3, and S-4 depict equations of linear discriminant scores for the test or training set for HPTLC fingerprints as well as for DART-MS fingerprints.

References

[1] J. Sherma, Anal. Chem. 82 (2010) 4895.

[2] C. Tistaert, B. Dejaegher, Y. Vander Heyden, Anal. Chim. Acta 690 (2011) 148.

[3] R. Pandey, S. Saraf, S. Saraf, Nat. Prod. Indian J. 3 (2007) 118.

[4] Z. Zeng, F. Chau, H. Chan, Ch. Cheung, T. Lau, Sh. Wei, D.K. Mok, Ch. Chan, Y. Liang, Chin. Med. 3 (2008) 9.

[5] B. Hemmateenejad, N. Mobaraki, F. Shakerizadeh-Shirazi, R. Miri, Analyst 135 (2010) 1747.

[6] R. Tian, P. Xie, H. Liu, J. Chromatogr. A 1216 (2009) 2150.

[7] K. Wuthold, G. Roos, U. Simmen, K.-A. Kovar, J. Planar Chromatogr. 16 (2003) 15.

[8] P. Taylor, A. Khatib, J. Liq. Chromatogr. Relat. Technol. 32 (2009) 2906.

[9] D. Milojković-Opsenica, P. Ristivojević, F. Andrić, J. Trifković, Chromatographia 76 (2013) 1239.

[10] E.S. Chernetsova, G.E. Morlock, I.A. Revelsky, Russ. Chem. Rev. 80 (2011) 235.

[11] J. Hajslova, T. Cajka, L. Vaclavik, Trends Anal. Chem. 30 (2011) 204.

[12] E. Werner, J.-F. Heilier, C. Ducruix, E. Ezan, C. Junot, J.-C. Tabet, J. Chromatogr. B 871 (2008) 143.

[13] M. Zhou, W. Guan, L.D. Walker, R. Mezencev, B.B. Benigno, A. Gray, F.M. Fernández, J.F. McDonald, Cancer Epidemiol. Biomarker Prevent. 19 (2010) 2262.

[14] T. Cajka, K. Riddellova, M. Tomaniova, J. Hajslova, Metabolomics 7 (2011) 500.

[15] L. Vaclavik, T. Cajka, V. Hrbek, J. Hajslova, Anal. Chim. Acta 645 (2009) 56.

[16] L. Vaclavik, V. Hrbek, T. Cajka, B.-A. Rohlik, P. Pipek, J. Hajslova, J. Agric. Food Chem. 59 (2011) 5919.

[17] S.M. Lee, H.-J. Kim, Y.P. Jang, Phytochem. Anal. 23 (2012) 508.

[18] K. Grob, J. Chromatogr. A 1150 (2007) 93.

[19] E.S. Chernetsova, A.S. Kovaleva, A.G. Koryakova, J. Anal. Chem. 65 (2010) 1420.

[20] G.E. Morlock, M.A. Vega-Herrera, J. Planar Chromatogr. 20 (2007) 411.

[21] G.E. Morlock, C. Oellig, J. AOAC Int. 92 (2009) 745.

[22] G.E. Morlock, W. Schwack, J. Planar Chromatogr. 20 (2007) 399.

[23] E.S. Chernetsova, I.A. Revelsky, G.E. Morlock, Anal. Bioanal. Chem. 401 (2011) 325.

[24] G.E. Morlock, W. Schwack, Trends Anal. Chem. 29 (2010) 1157.

[25] N. Kunz, I. Scholl, A. Schroeder, G. Morlock, Planar chromatography fingerprint of German propolis. 58th Annual convention of the Association of the German Bee Research Institutes, Berlin, Germany, March 29-31, 2011, Poster P 21.

[26] E.S. Chernetsova, P.O. Bochkov, M.V. Ovcharov, S.S. Zhokhov, R.A. Abramovich, Drug Test. Anal. 2 (2010) 292.

[27] C. Sârbu, A.C. Mot, Talanta 85 (2011) 1112.

[28] O.L. Ogegbo, S. Eyob, S. Parmar, Z.-T. Wang, S.W.A. Bligh, Anal. Methods 4 (2012) 2522.

[29] C.-J. Xua, Y.-Z. Liang, F.-T. Chauc, Y. Vander Heyden, J.Chromatogr. A 1134 (2006) 253.

[30] Y.-Z. Liang, P. Xie, K. Chan, J. Chromatogr. B 812 (2004) 53.

[31] L. Giansante, D. Di Vincenzo, G. Bianchi, J. Sci. Food Agric. 83 (2003) 905.

[32] E.S. Chernetsova, M. Bromirski, O. Scheibner, G.E. Morlock, Anal. Bioanal. Chem. 403 (2012) 2859.

[33] G.E. Morlock, E.S. Chernetsova, A. Schroeder, N. Kunz, I. Scholl, New HPTLC method for analysis of flavonoids and phenolic compounds in propolis. International Symposium

for High-Performance Thin-Layer Chromatography, Basel, Switzerland, July 06-08, 2011, Poster P-8j.

[34] M. Chudzinska, D. Baralkiewicz, Food Chem. Toxicol. 48 (2010) 284.

[35] S. Sun, B. Guo, Y. We, M. Fan, Food Chem. 124 (2011) 1151.

[36] F. Pellati, G. Orlandinia, D. Pinetti, S. Benvenutia, J. Pharm. Biomed. Anal. 55 (2011) 934.

[37] S.I. Falcão, N. Vale, P. Gomes, M.R.M. Domingues, C. Freire, S.M. Cardosoe, M. Vilas-Boasa, Phytochem. Anal. 24 (2013) 309.

[38] T. Cajka, K. Riddellova, M. Tomaniova, J. Hajslova, J. Chromatogr. A 1217 (2010) 4195.

[39] I. Stanimirova, B. Ustun, T. Cajka, K. Riddelova, J. Hajslova, L.M.C. Buydens, B. Walczak, Food Chem. 118 (2010) 171.

[40] A.C.H.F. Sawaya, D.M. Tomazela, I.B.S. Cunha, V.S. Bankova, M.C. Marcucci, A.R. Custodio, M.N. Eberlin, Analyst 129 (2004) 739.

[41] D. Arráez-Román, A.M. Gómez-Caravaca, M. Gómez-Romero, A. Segura-Carretero, A. Fernández-Gutiérrez, J. Pharm. Biomed. Anal. 41 (2006) 1648.

[42] M. Biesaga, K. Pyrzynska, J. Chromatogr. A 1216 (2009) 6620.

Table 1

HPTLC: LDA classification of propolis extracts from the training set of orange and blue types.

Classification results[a,c]					
		Class	Predicted Group Membership		Total
			Blue type	Orange type	
Original	Count	Blue type	26	1	27
		Orange type	1	31	32
	%	Blue type	96.3	3.7	100.0
		Orange type	3.1	96.9	100.0
Cross-validated[b]	Count	Blue type	26	1	27
		Orange type	4	28	32
	%	Blue type	96.3	3.7	100.0
		Orange type	12.5	87.5	100.0

a. 96.6 % of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 91.5 % of cross-validated grouped cases correctly classified.

Table 2

Classification and interpretation of characteristic signals in DART mass spectra of two propolis types found by principle component analysis

#	Variable, *m/z*	Impact of variable on PC	Assumed marker compounds*	Characteristic for propolis type
1	163.1	positive (PC1)	coumaric acid	blue and orange
2	242.1	negative (PC2)	unknown phenolic compound	blue
3	253.1	positive (PC1), positive (PC2)	chrysin or methoxyflavanone	blue and orange
4	255.1	positive (PC2)	pinocembrin or liquiritigenin	orange
5	269.1	positive (PC2)	galangin, apigenin, pinostrobin or benzyl caffeate	blue and orange
6	271.1	positive (PC2)	naringenin or pinobanksin	orange
7	285.1	positive (PC2)	kaempferol, luteolin, pinobanksin-5-methyl-ether	blue
8	327.1	positive (PC1)	pinobanksin-5-methylether-3-O-acetate or pinobanksin-3-O-propionate	blue
9	332.1	positive (PC2)	unknown phenolic compound	blue and orange
10	343.1	positive (PC1)	unknown phenolic compound	blue
11	417.1	positive (PC1)	unknown phenolic compound	blue
12	449.1	positive (PC2)	unknown phenolic compound	blue and orange
13	517.2	negative (PC1)	glycosides or phenolic dimer	orange
14	541.1	negative (PC1)	glycosides or phenolic dimer	orange
15	583.1	negative (PC1)	glycosides or phenolic dimer	orange

*Deprotonated molecules were considered as source for respective *m/z* signals.

460

461 **Table 3**

462 DART-MS: LDA classification of propolis extracts from the training set of orange and blue

463 types.

Classification Results[a,c]					
		Class	Predicted Group Membership		Total
			Blue type	Orange type	
Original	Count	Blue type	24	2	26
		Orange type	0	33	33
	%	Blue type	92.3	7.7	100.0
		Orange type	0	100.0	100.0
Cross-validated[b]	Count	Blue type	24	2	26
		Orange type	1	32	33
	%	Blue type	92.3	7.7	100.0
		Orange type	3.0	97.0	100.0

a. 96.6 % of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 94.9 % of cross-validated grouped cases correctly classified.

465

List of figures

Fig. 1. Clustering on a 2D PC-score (A) and loading plot (B) based on zone intensities of the HPTLC image (see Fig. 2).

Fig. 2. HPTLC fingerprint [25]: Characteristical zone pattern for blue and orange types of propolis samples with mass signal assignment.

Fig. 3. Dendrograms of propolis samples (A) and variables (B) based on the HPTLC fingerprints (see Fig. 2).

Fig. 4. Linear discriminant scores for orange and blue propolis types (A) and mutually the additional set of samples of foreign and mixed orange-blue types (B) based on the HPTLC fingerprints.

Fig. 5. DART mass spectra (T_{DART} = 300 °C) exemplarily shown for blue (A, C) and orange propolis types (B, D).

Fig. 6. Clustering on a 2D PC-score (A) and loading plot (B) based on selected DART mass signals (see Fig. 5).

Fig. 7. Dendrograms of propolis samples (A) and variables (B) based on selected DART mass signals (see Fig. 5).

Fig. 8. Linear discriminant scores for orange and blue propolis types (A) and mutually the additional set of samples of foreign and mixed orange-blue types (B) based on selected DART mass signals.

Highlights

- Novel combination of two fast fingerprint techniques provided complementary information
- Multivariate data analysis improved the differentiation of natural products
- HPTLC and DART-MS for fast pattern recognition and categorization of samples by chemometrics
- HPTLC provided information on polarity, functional groups and spectral properties
- DIP-it DART-MS furnished information on elemental formulae of principal components

Figure 1

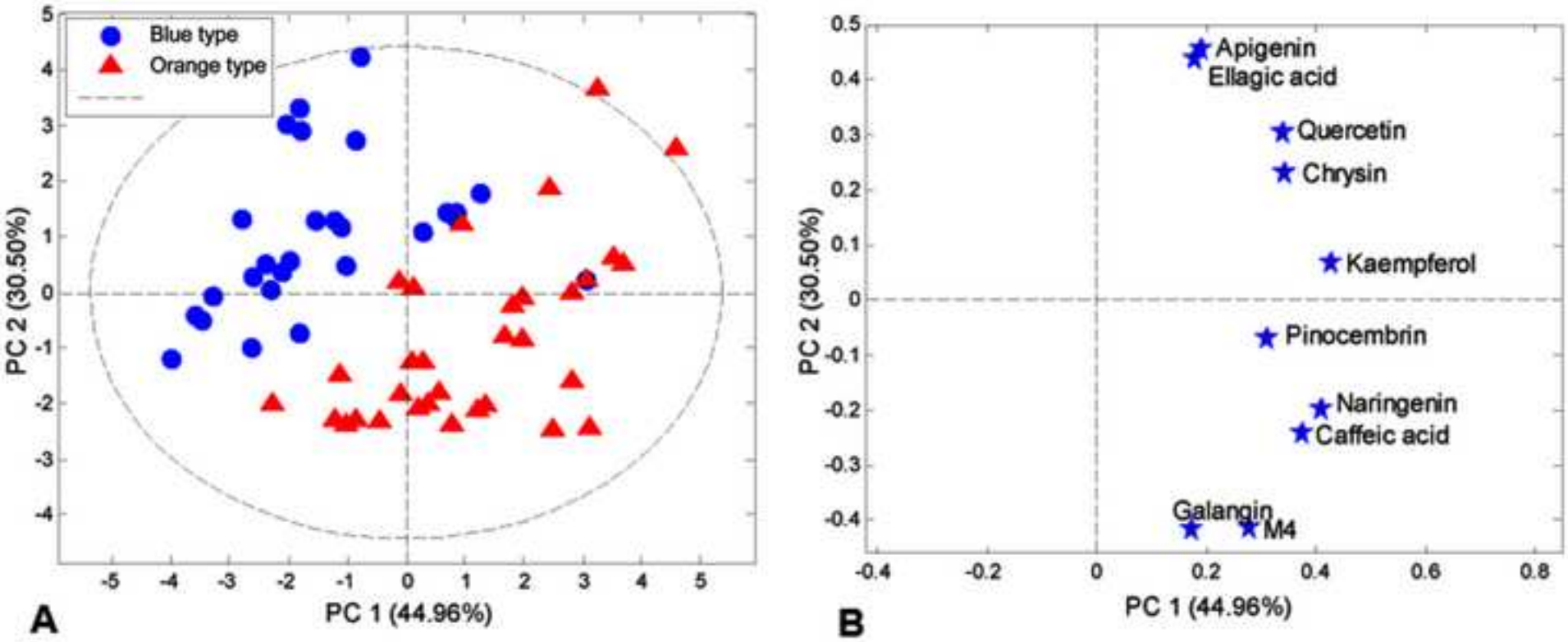

Blue type
Orange type
PC 2 (30.50%)
PC 1 (44.96%)
A
Apigenin
Ellagic acid
Quercetin
Chrysin
Kaempferol
Pinocembrin
Naringenin
Caffeic acid
Galangin
M4
PC 2 (30.50%)
PC 1 (44.96%)
B

Figure 2

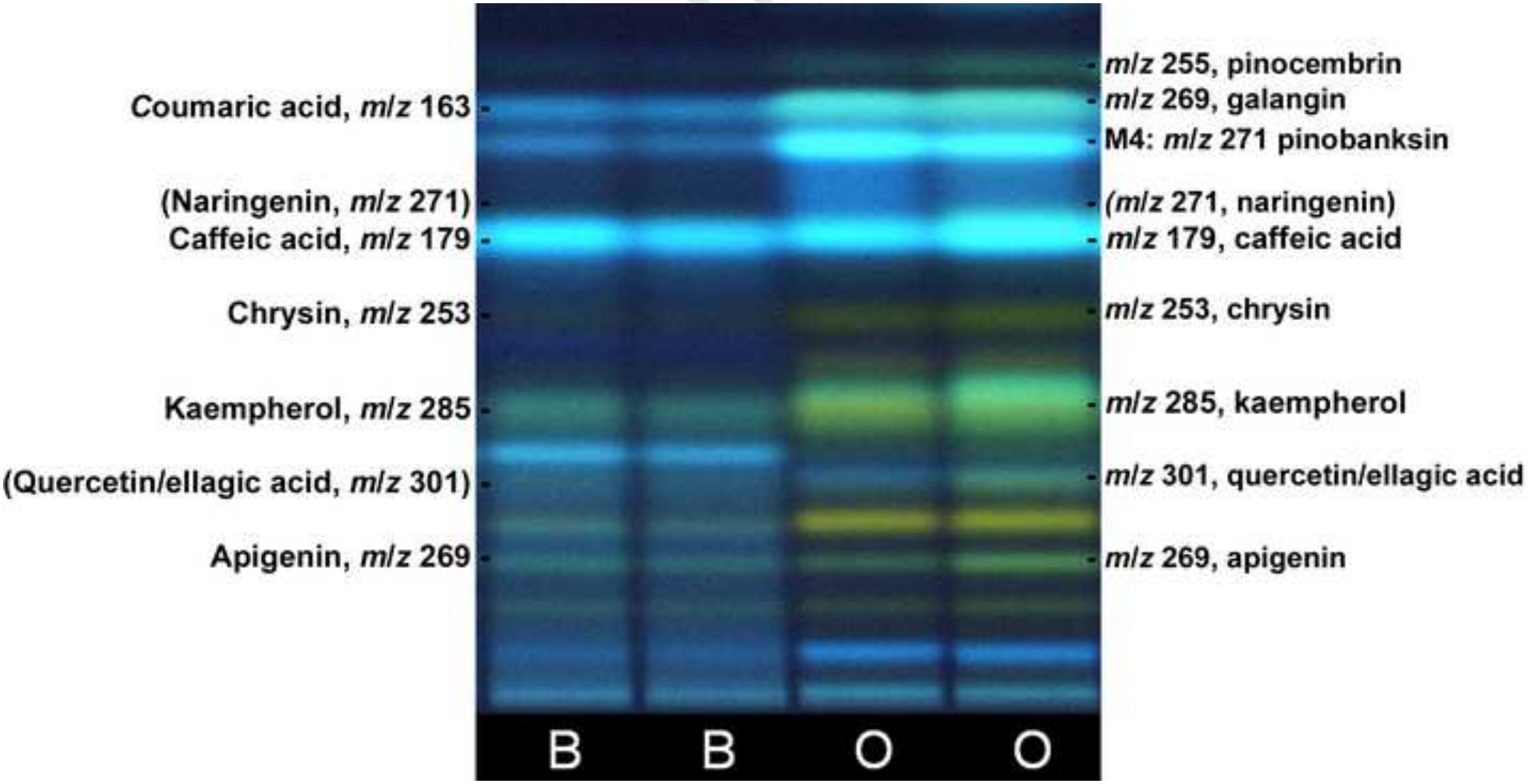

Coumaric acid, m/z 163
(Naringenin, m/z 271)
Caffeic acid, m/z 179
Chrysin, m/z 253
Kaempherol, m/z 285
(Quercetin/ellagic acid, m/z 301)
Apigenin, m/z 269
m/z 255, pinocembrin
m/z 269, galangin
M4: m/z 271 pinobanksin
(m/z 271, naringenin)
m/z 179, caffeic acid
m/z 253, chrysin
m/z 285, kaempherol
m/z 301, quercetin/ellagic acid
m/z 269, apigenin
B
B
O
O

Figure 3

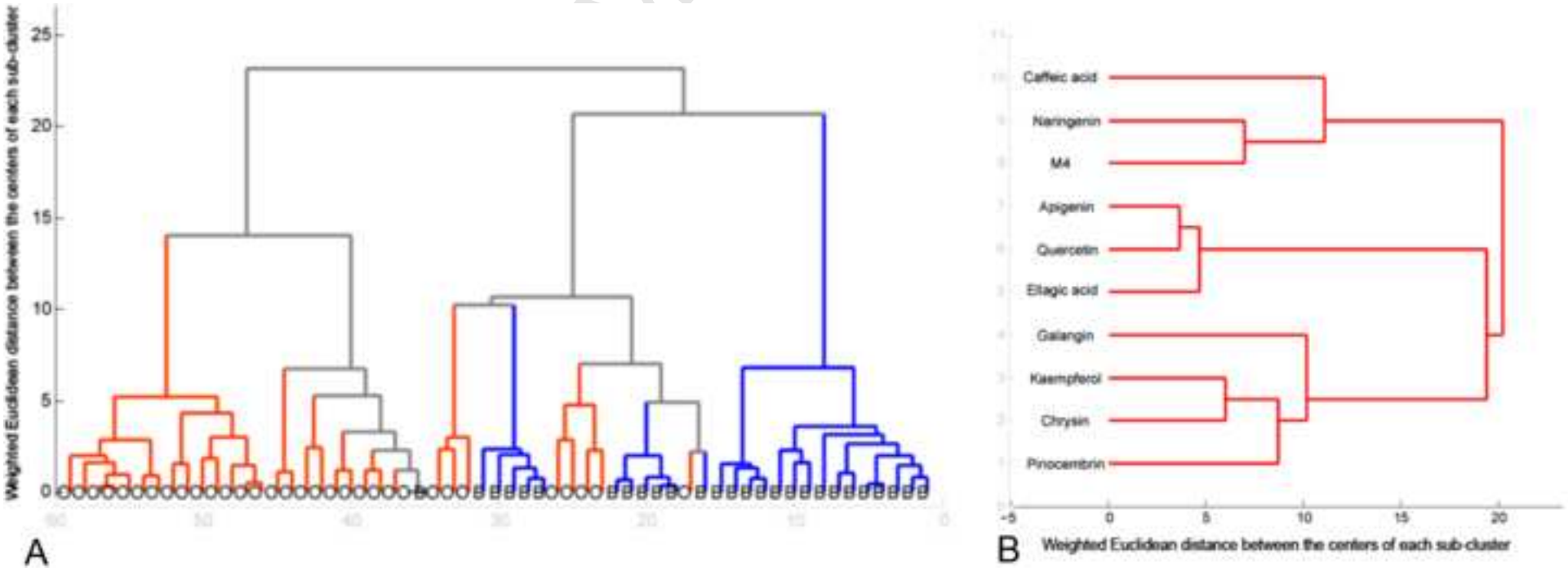

Weighted Euclidean distance between the centers of each sub-cluster
25
20
15
10
5
0
A
Caffeic acid
Naringenin
M4
Apigenin
Quercetin
Ellagic acid
Galangin
Kaempferol
Chrysin
Pinocembrin
-5
0
5
10
15
20
B
Weighted Euclidean distance between the centers of each sub-cluster

Figure 4

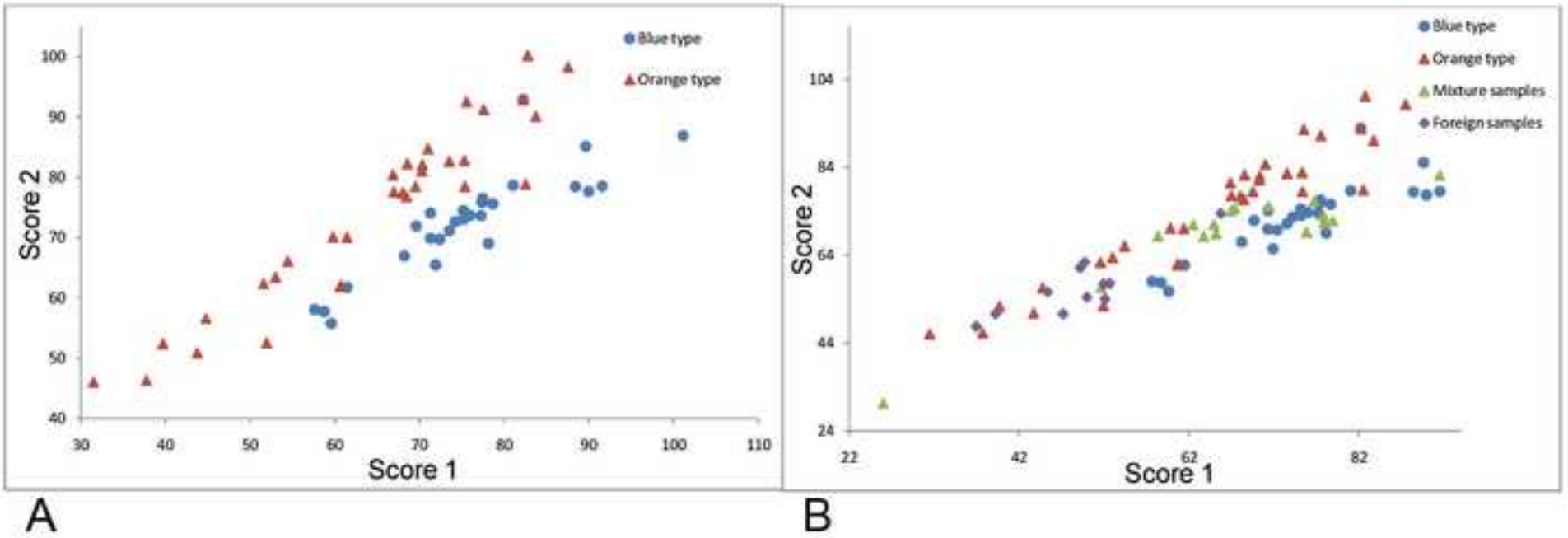

Blue type
Orange type
Score 2
Score 1
100
90
80
70
60
50
40
30
40
50
60
70
80
90
100
110
A
Blue type
Orange type
Mixture samples
Foreign samples
Score 2
Score 1
104
84
64
44
24
22
42
62
82
B

Figure 5

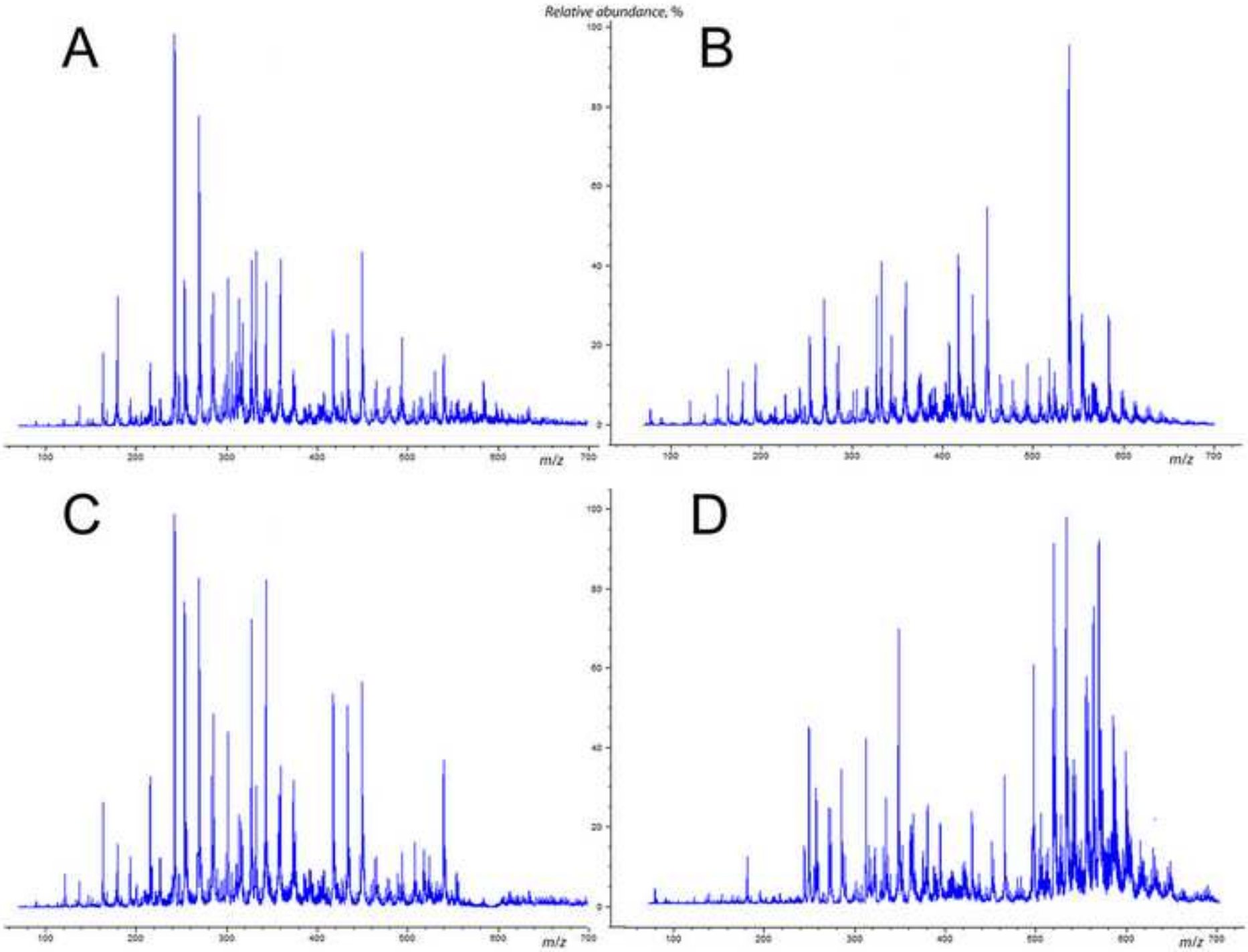
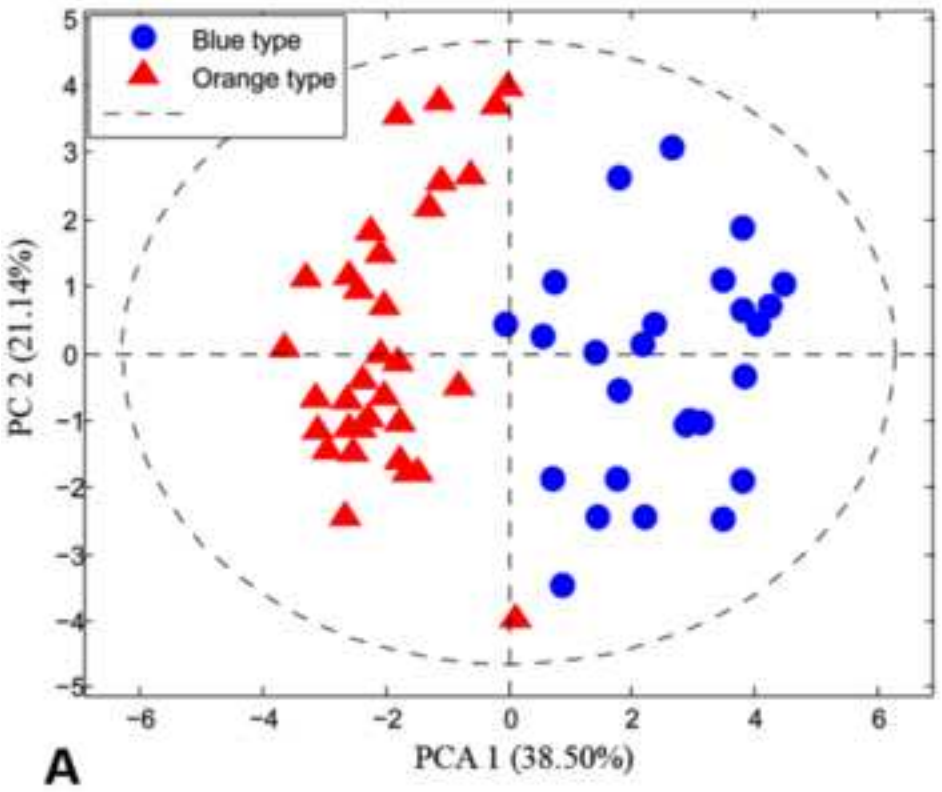


Figure 6


Blue type
Orange type
PC 2 (21.14%)
PCA 1 (38.50%)
A

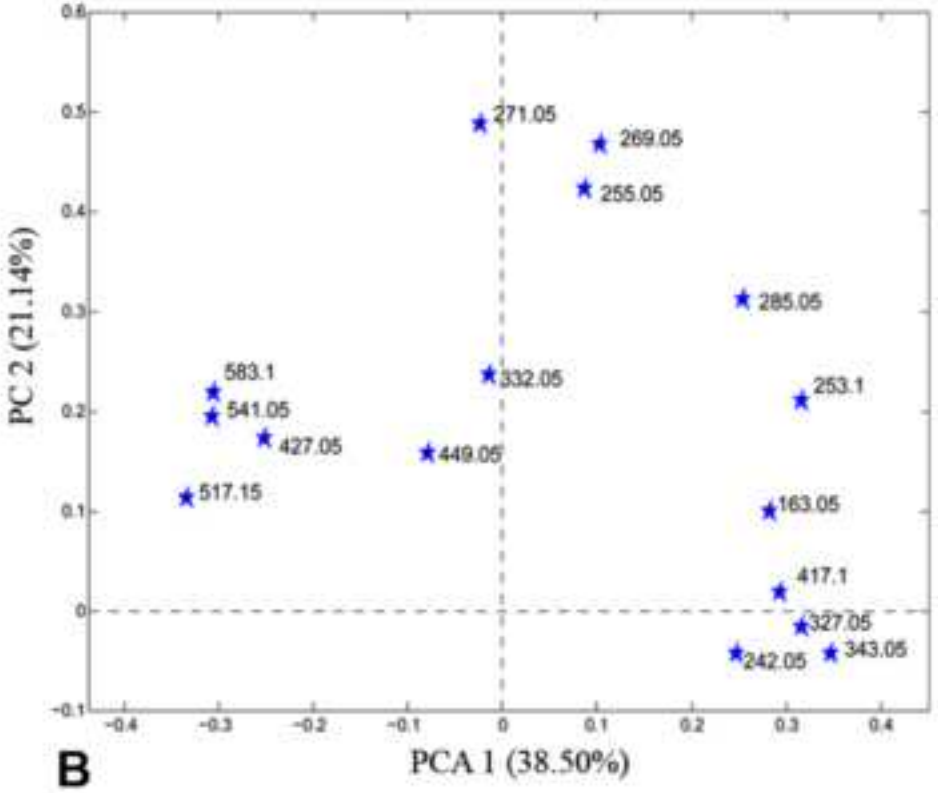

271.05
269.05
255.05
285.05
583.1
332.05
253.1
541.05
427.05
449.05
517.15
163.05
417.1
327.05
242.05
343.05
PC 2 (21.14%)
PCA 1 (38.50%)
B

Figure 7

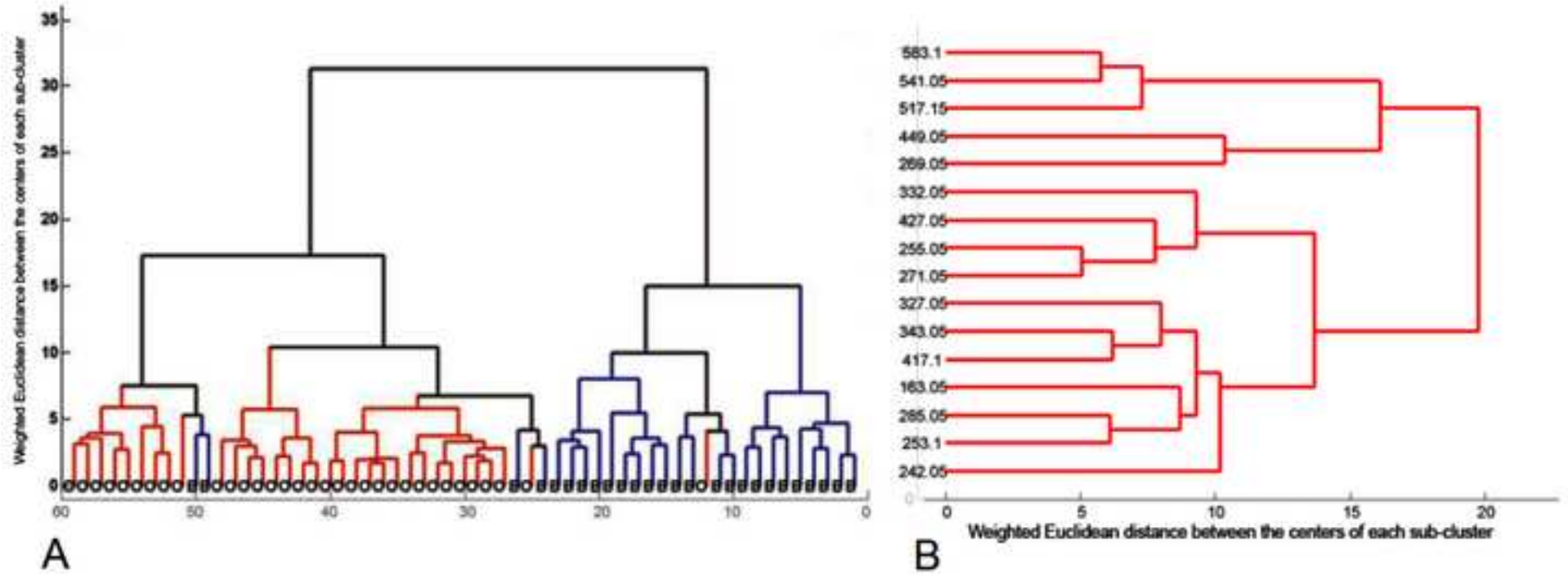
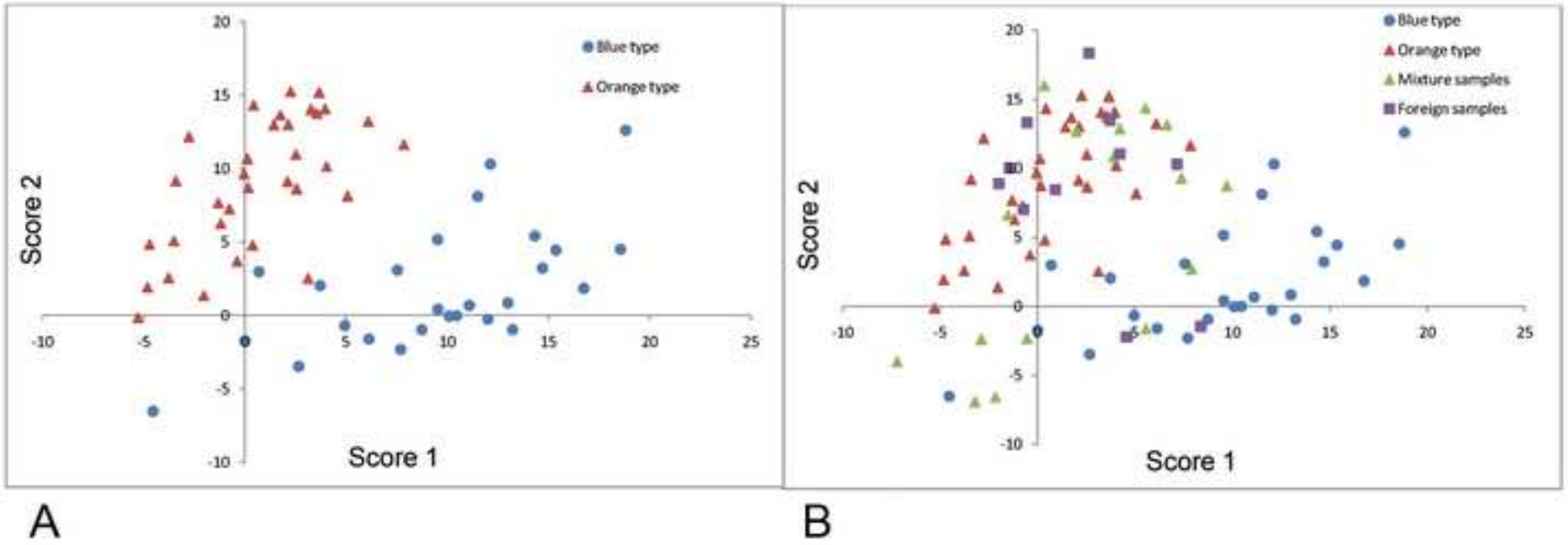

Weighted Euclidean distance between the centers of each sub-cluster
35
30
25
20
15
10
5
0
60
50
40
30
20
10
0
A
583.1
541.05
517.15
449.05
269.05
332.05
427.05
255.05
271.05
327.05
343.05
417.1
163.05
285.05
253.1
242.05
0
5
10
15
20
Weighted Euclidean distance between the centers of each sub-cluster
B

Figure 8


Blue type
Orange type
Score 2
Score 1
20
15
10
5
0
-5
-10
-10
-5
5
10
15
20
25
A
Blue type
Orange type
Mixture samples
Foreign samples
Score 2
Score 1
B