Analytical Methods

PAPER

View Article Online
View Journal | View Issue

Cite this: *Anal. Methods*, 2014, **6**, 3013

In quest for chemomarkers to classify Taiwanese teas†

Ewelina Czerniec-Michalik, Kai-Ta Hsieh and Pawel L. Urban*

Received 12th November 2013
Accepted 30th January 2014

DOI: 10.1039/c3ay42009d

www.rsc.org/methods

It is of great economical and practical significance to distinguish between different classes, brands, and places of origin of commercial tea. Herein, an analytical method has been developed to characterize samples of oriental teas (with an emphasis on Taiwanese teas). It is based on the implementation of liquid chromatography coupled with mass spectrometry along with chemometric data treatment. Prior to chemometric classification, datasets (for over 80 samples) were subjected to an exhaustive pre-treatment procedure in order to select potential chemomarker signals according to their variability within pre-defined groups of samples. Statistical analysis further revealed differences in the composition of the teas originating from plantations in Taiwan and other countries as well as the teas belonging to the major classes defined by the degree of fermentation. The greatest differences relate to the content of caffeine and epi/catechin represented by the mass spectrometric signals at the *m*/*z* 195.0 and 291.0, respectively. The proposed analytical workflow provides a convenient tool to reveal differences in the chemical composition of multi-component real samples such as plant tissue extracts.

1 Introduction

Tea is the second most frequently consumed beverage in the world – right next to water.[1] Tea infusions are prepared from processed leaves and flowers of tea plant (*Camellia sinensis*). Properties of tea are characteristic of the origin of plant material, as well as its cultivar. They vary according to the climatic conditions of the cultivation area, the time of harvest as well as the production process and storage conditions. Sensory characteristics of tea infusions are closely associated with the chemical components present in tea leaves. Three types of tea – black (traditional Mandarin: 紅茶; pinyin: *hóng chá*), green (綠茶; *lǜ chá*) and oolong (烏龍; *wū lóng*) – defined on the basis of the degree of fermentation – constitute the most important classes.[1] Because of high market demand, production of tea in the Far East is highly competitive. Steps of tea production are meticulously controlled by manufacturers, and remain their trade secret. Cases of unethical behaviour of sellers (adulteration) have been revealed.[2] Nonetheless, for non-professionals, it is hard to judge on the quality of tea leaves. Application of appropriate reliable testing and certification methods can assist the enforcement of quality guidelines.

Differentiation of tea varieties according to the harvesting seasons and the production areas can be performed by using near infrared spectroscopy,[3] laser-induced fluorescence,[4] inductively coupled plasma (ICP) mass spectrometry (MS) and gene expression profiling methods.[5] In other studies, a Taiwanese green tea, GABA tea (containing γ-aminobutyric acid) as well as tea beverages were characterized using classical and chromatographic methods of analysis.[6,7] Contents of ascorbic acid, caffeine and phenolic compounds were determined, although the applied procedures were highly time consuming and laborious while they provided information about the content of specific compounds only. A number of simple and reliable chromatographic methods have been developed for the quality control of tea and tea beverages. For example, seven bioactive components in oolong tea from different production areas (Guangdong and Fujianin, China, as well as Taiwan) were characterized using a high-performance liquid chromatography method with UV absorption detection.[8]

While only complex multi-platform analytical approaches could enable comprehensive characterization of the chemical composition of tea, for practical reasons, it is appealing to choose a single robust and fast analytical platform, which could provide datasets of sufficient quality to enable classification of oriental teas. Due to its speed, efficiency, and reliability in numerous metabolomics-related analyses, liquid chromatography (LC) hyphenated with MS seemed to be a suitable candidate technique for this task. However, similar to many studies in systems biology, collecting raw data is only the first step in the search for biochemically significant marker molecules. The development of accurate and sensitive analytical methods to enforce adequate quality assurance procedures in tea trade is a challenge for analysts. Constant control is of

Department of Applied Chemistry, National Chiao Tung University, Hsinchu 300, Taiwan. E-mail: plurban@nctu.edu.tw

† Electronic supplementary information (ESI) available: Additional tables and figures. Table S1 (Excel file) includes the list of samples along with general information. More details on the samples (only for scientific purposes) can be obtained on request. See DOI: 10.1039/c3ay42009d

highest importance when it comes to the maintenance of high production standards as well as the enforcement of trade regulations. Therefore, the goal of this work was to develop a methodology suitable for the discrimination of tea kinds – black, green and oolong, as well as for the differentiation of Taiwanese tea brands from the tea produced in other countries. It was hypothesized that an efficient interpretation of LC-MS datasets representing a multitude of chemical features could lead to the selection of several putative chemomarkers of tea classes and origin areas.

2. Materials and methods

2.1. Materials and samples

The samples of tea leaves (*Camellia sinensis*) produced in Taiwan ($n = 64$), Sri Lanka ($n = 7$), United Kingdom ($n = 6$), South Korea ($n = 1$), Indonesia ($n = 1$), Italy ($n = 1$), Japan ($n = 1$), and Singapore ($n = 1$) were used in this study. The choice of the samples was quasi-random, and mostly dictated by the availability of the teas in local shops in Hsinchu City (Taiwan). It was assumed that – for the initial testing of the proposed method – it would be desirable to include a broad range of teas (from less expensive to more costly Taiwanese brands; as well as the brands marketed in other countries in different continents; Table S1†). Samples used in the statistical analysis were further selected (for example, samples with flavor additives were excluded in some analyses).

The raw samples were initially extracted with deionized water at 95 °C for 20 min (to obtain extracts with similar chemical compositions to those of tea infusions). The extracts were then centrifuged, and the supernatant was filtered using 0.2 μm polytetrafluoroethylene (PTFE) syringe filters (Pall Corporation, Ann Arbor, USA). Samples were freshly prepared prior to the analysis. Methanol for LC-MS (≥99.9%, Chromasolv, Fluka, Sigma-Aldrich, St Louis, USA) LC-MS grade water (Chromasolv, Fluka, Sigma-Aldrich), acetic acid (49–51%, Sigma-Aldrich) and formic acid (49–51%, Fluka, Sigma-Aldrich) were used for the preparation of the mobile phase and as extracting agents.

Identification and quantification of caffeine and catechin species were performed using a certified reference material containing: (−)-epigallocatechin 3-gallate (100.0 ± 0.8 μg mL^{-1}), (+)-catechin (99.9 ± 0.8 μg mL^{-1}), (−)-epicatechin (100.0 ± 0.8 μg mL^{-1}), (−)-epicatechin 3-gallate (100.1 ± 0.8 μg mL^{-1}), (−)-gallocatechin (99.9 ± 0.8 μg mL^{-1}), (−)-gallocatechin 3-gallate (100.0 ± 0.8 μg mL^{-1}), (−)-catechin 3-gallate (100.1 ± 0.8 μg mL^{-1}) and caffeine (100.1 ± 0.8 μg mL^{-1}) dissolved in the mixture of acetonitrile and 5% 1 M hydrochloric acid in water (80 : 20, v/v) (Sigma-Aldrich).

2.2. Liquid chromatography conditions

A Dionex UltiMate 3000 ultra-high performance liquid chromatography system (Thermo Fisher Scientific Inc, Waltham, USA) equipped with a Hypersil Gold AQ C18 column (150 mm × 1 mm I.D., 3 μm) was used for the separation. The column temperature was set to 30 °C, while the flow rate of the mobile phase was set to 20 μL min^{-1}. Chromatographic conditions were optimized while operating the LC with an active pixel sensor-based UV detector (Paraytec Ltd., York, UK), incorporating a pulse xenon lamp and a 200–210 nm bandpass filter. Mobile phases were composed of (A) 0.2% acetic acid (for UV detection) or 0.2% formic acid (for MS detection) in deionized water, and (B) methanol. The gradient elution program was as follows: 15–80% (B) from 0 to 45 min, 80–15% (B) from 45 to 46 min, followed by a 5 min equilibration stage with 15% (B). It became evident that small changes of the separation conditions did not greatly affect the separation efficiency of the method (Fig. S1†).

2.3. Mass spectrometry

The mass spectra were collected using an Amazon Speed ion-trap mass spectrometer equipped with an electrospray ionization (ESI) source (Bruker Daltonics, Bremen, Germany), operating in the UltraScan positive-ion mode. Since all the effluent was directed to the ESI nebulizer, the ion source was operated at a flow rate of 20 μL min^{-1}. The voltage applied to the MS transfer capillary was kept at +4 kV and end plate offset voltage was kept at 300 V. Drying gas (nitrogen) was heated up to 250 °C, and pumped with a flow rate of 6 L min^{-1}. Nebulising gas pressure was set to 15 psi (~103 kPa). The m/z range of the mass analyzer was set to 100–600 u e^{-1}. The accumulation time was set to 10 ms. Each spectrum acquired was the sum of 10 scans. The ion source parameters for MS/MS analysis were kept the same as for LC-MS analysis. Fragmentation was performed using scheduled precursor lists created according to the results of the principal component analysis (PCA). The number of precursor ions was set to 1. The time tolerance was ±30 s while the m/z tolerance was ±0.5 u e^{-1}.

2.4. Data treatment

The first stage of data treatment was conducted using ProfileAnalysis software (version 2.1; Bruker Daltonics). All the 3D data (time-m/z-intensity), obtained during the analysis of tea samples, were aligned, and processed using the bucketing method. Retention times (RTs) were aligned according to the shifting vector algorithm described elsewhere.[9] Such an alignment was necessary because of run-to-run variability, which would affect the statistical classification.

Transformation of complex 3D data (distributed in the time–m/z–intensity space) to two dimensions was performed using a mapping method. This resulted in the formation of a data matrix, where intensity values for each sample were distributed into buckets (time–m/z co-ordinates).[10] The chromatographic time range used for the treatment of each dataset was 240–2700 s, the m/z range was 100–600 u e^{-1}, the bucket size was 60 s × 1 u e^{-1}, and kernelizing parameters were set to 30 s and 0.2 u e^{-1}. The signal-to-noise threshold was set to the value of 10, while the minimum number of spectra, in which ions of a compound should be detected to be considered in the bucketing, was set to 10. In addition, spectrum background subtraction was applied.

Bucket matrices were obtained using three ways of analysis: (i) general bucketing, (ii) group bucketing, and (iii) a combination of

(i) and (ii) (Fig. 1). In the first way of statistical treatment – general bucketing – all samples were initially bucketed together using the commercial software from Bruker Daltonics. Subsequently, the obtained bucket values were divided into three groups according to their kind – representing black, green and oolong tea – or two groups – corresponding to the country of origin (Taiwan/other country) within one kind (black or green tea). The relative standard deviation (RSD) describing the sample-to-sample variability within each group was calculated for every bucket. Thirty buckets with the lowest RSD in each group were selected and used for statistical analysis. The number of buckets selected for analysis is arbitrary; this number has been chosen following a sequence of preliminary tests.

In the second way of statistical treatment – group bucketing – samples were first divided into three groups – corresponding to black, green and oolong tea. Subsequently, bucketing was conducted. This was followed by the calculation of RSD values for each bucket. The obtained RSD values were ordered from the lowest to the highest one, and thirty buckets from each group were selected for further analysis in a similar manner as described above. Note that in both cases ((i) and (ii)), the final number of buckets used in analysis was greater than 30 because only some of the buckets selected for each group overlapped.

The third way of statistical data treatment was a combination of buckets chosen by group bucketing (ii) and values from general bucketing (i). In the case of all three methods ((i), (ii), and (iii)), the values of the buckets were normalized taking into account the sample mass (since it was impossible to set the weight of every sample to the exact value of 100.00 mg), and standardized by dividing each value by the standard deviation of each bucket using the MATLAB software (version 7.14; MathWorks, Natick, Massachusetts, USA).

After this pre-treatment, the datasets were used in PCA analysis performed in MATLAB. The obtained scores plots were overlaid with vectors corresponding to the bucket variables that significantly influenced the distribution of the data points.

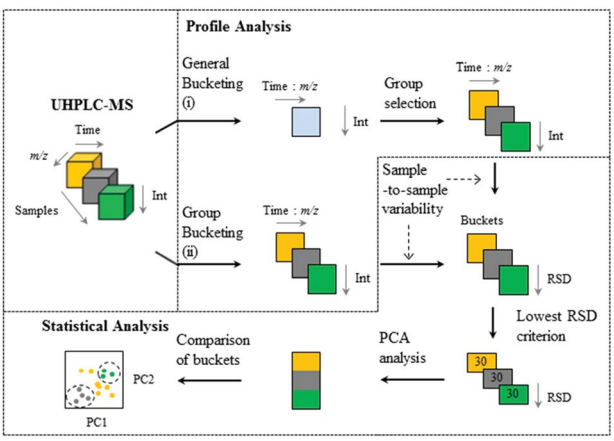

Fig. 1 Data processing workflow. Conversion of the raw LC-MS data to a 2D bucket matrix. Selection of buckets according to their variability within samples followed by multivariate analysis.

3. Results and discussion

3.1. LC-MS method

The run time of the optimized LC method was 50 min, which was sufficient for the separation of tea components adequate for ESI-MS, and subsequent classification of the composition profiles. No carryover effect was observed. The implementation of a micro-scale column and capillary connectors reducing dwell-volume during analysis makes this chromatographic methodology relatively fast and solvent-efficient (~1 mL per analysis). After initial processing of the raw LC-MS data, the preliminary identification of compounds could be performed by analyzing 2D spectral/chromatographic maps (Fig. 2). Using this visualization method, one can easily observe signals at m/z values characteristic of each tea sample. More importantly, the developed LC-MS method resolved highly abundant catechin epimers, such as gallocatechin, recorded at the m/z 307.1 u e^{-1}, and RTs 13.9 and 20 min.

3.2. Chemometric and statistical analysis

PCA is a popular statistical method which enables the observation of groupings within multivariate datasets. A data matrix

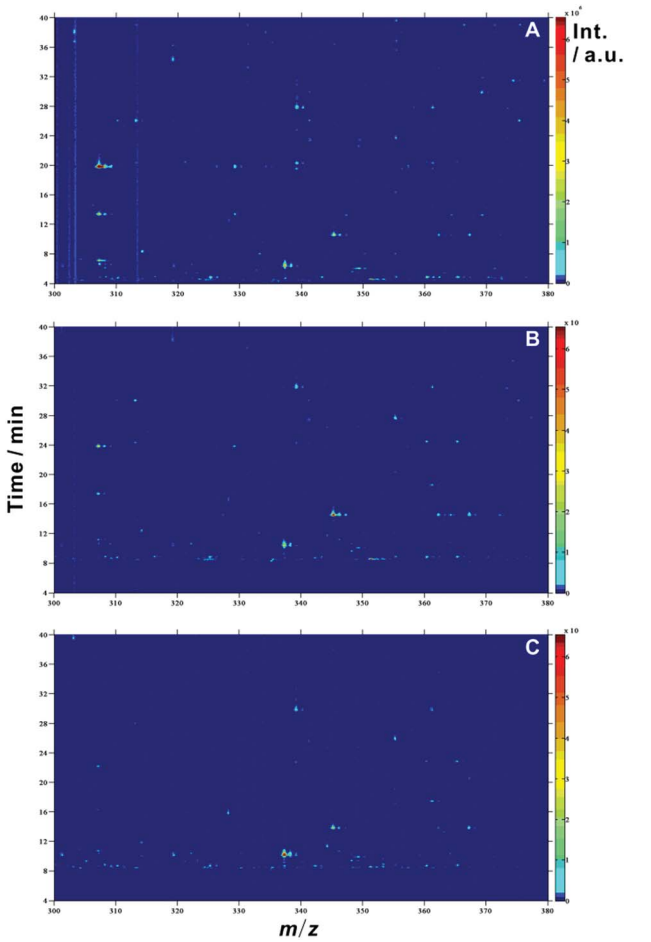

Fig. 2 Examples of 2D data (LC-MS). A – green tea (#1), B – oolong tea (#2), and C – black tea (#15).

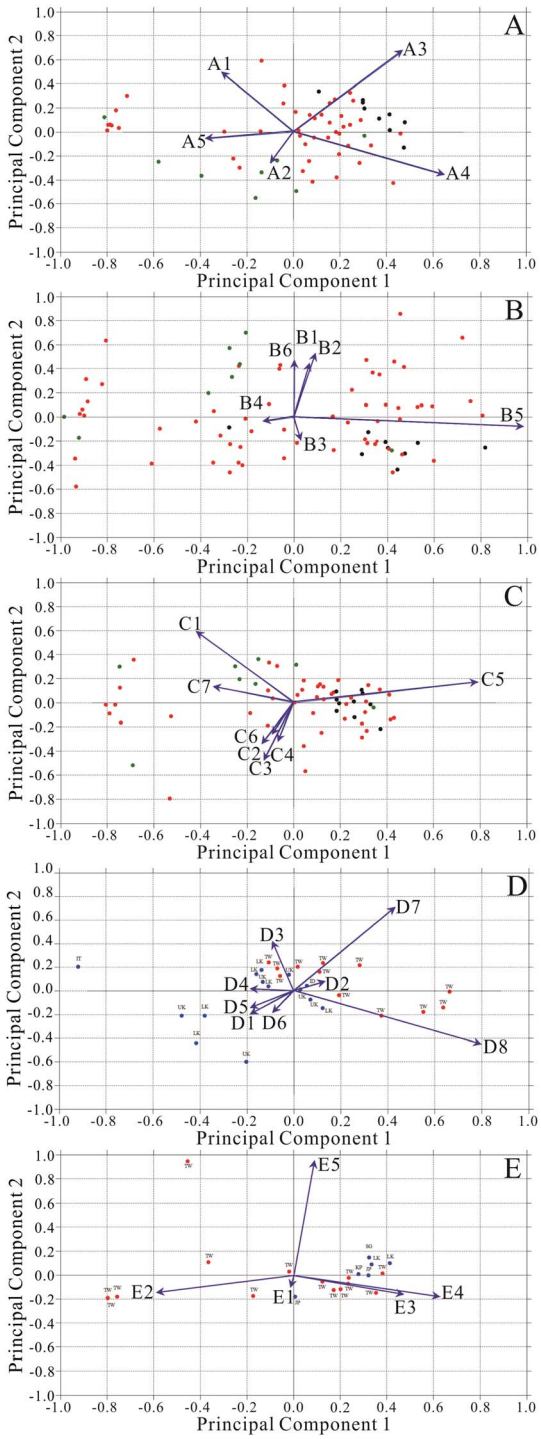

Fig. 3 PCA scores plots of Taiwanese tea: A – group bucketing, B – general bucketing and group bucket selection, and C – general bucketing (for a 3D plot, see Fig. S2†). PCA scores plots of D – black tea and E – green tea were created based on general bucketing. Markers in (A–C) red – oolong tea; green – green tea; and black – black tea. Markers in (D,E) red – Taiwanese tea; and blue – other tea. Note well: some markers overlap. Countries of origin: TW – Taiwan, JP – Japan, KP – South Korea, SG – Singapore, LK – Sri Lanka, UK – United Kingdom, IT – Italy, and ID – Indonesia. The displayed vectors correspond to the buckets with a high sample-to-sample variability. See Table 1 for the explanation of the symbols corresponding to individual buckets.

containing variables obtained for every sample is converted into a matrix of principal components describing the maximum variation within the data. The result of PCA can be represented by the scores plot, which shows distribution of samples (observations) in the dataset or loading plot – the graph of coefficients by which the input variables have to be multiplied to obtain the corresponding principal component.[11] In this study, the enhanced scatter plot (biplot) was applied. Biplot points correspond to the observations while vectors represent the variable coefficients of the principal components.[12] Vectors pointing in the same direction have similar response profiles.[13]

In Fig. 3, observations represent tea samples and the vectors represent buckets corresponding to chemical compounds present in tea. The impact of each variable on the statistical model can be illustrated with the vectors displayed in the scores plot. The differences between green, oolong and black teas (Fig. 3A–C and S2†), as well as between Taiwanese teas and the foreign teas (Fig. 3D and E) can be seen in the corresponding scores plots. The data points are separated to some extent only. For instance, outliers – falling into the zones occupied by the data points pertaining to different groups – hinder classification of samples. Thus, further improvements in the data curation strategy should be sought in future studies.

Interestingly, statistical treatment of the raw data (without the processing shown in Fig. 1) did not lead to satisfactory results (Fig. S3†). This negative result is explained with the high complexity and similarity of the composition of all investigated samples. Due to the large number of factors influencing the composition of processed tea (*e.g.* plant genotype, location of plantation site, climate and weather, time of harvest, fermentation, baking, storage, and any other steps taken by the manufacturer), the values of all buckets vary a lot among the samples of teas – within the classes and globally. The multitude of "noisy buckets" decreases the chances for successful separation of tea according to the subjective classification (*e.g.* oolong/green/black or Taiwan/other country). Picking the variables (buckets) according to the lowest-RSD criterion (*cf.* Fig. 1), while considering the classes of interest, helps to differentiate teas according to chosen criteria and select potential chemomarkers.

Application of various ways of data treatment (*cf.* Section 2.4) revealed potential chemomarkers of different classes of tea. Buckets with the highest impact on scores distribution – as judged based on the length of the corresponding vectors in Fig. 3 – are listed in Table 1. For example, the bucket denoted as 1350 s:194.5 u e^{-1} corresponds to caffeine (m/z 194.9 u e^{-1}, RT 23.3 min). The relative content of caffeine can be the differentiating factor for oolong and black tea groups, as well as of Taiwanese tea and teas of other origin. Higher concentration values of caffeine can be observed in black teas, when compared to green and oolong teas. Curiously, Taiwanese teas contain more caffeine in relation to the teas from other countries included in this study. It is pleasing to note that one can observe a clear separation between the caffeine-free black tea sample (Deteinato, Bologna, Italy) and other black teas (Fig. 3D). Another compound which particularly contributes to the resolution of green and oolong teas is one of the

Table 1 Variables (buckets) – represented by vectors – which contribute most to the separation of data points in the scores plots of PCA (Fig. 3)

Symbol	Bucket time/s : m/z/u e^{-1}	Analyte ion m/z/u e^{-1}
D1	270 : 138.5	138.9
D2	270 : 192.5	192.7
C2	330 : 192.5	192.7
B1 and D3	270 : 351.5	351.1
A1	270 : 381.5	381.4
D4	270 : 385.5	384.9
B3 and C3	330 : 385.5	384.9
B2 and C1	330 : 174.5	174.9
A2 and D6	390 : 174.5	174.9
D5	330 : 214.5	214.8
C4	330 : 407.5	407.1
E1	390 : 157.5	158.0
E2	390 : 337.5	337.3
A3, B4 and D7	1290 : 194.5	194.9
A4, B5, C5, D8 and E3	1350 : 194.5	194.9
E4	1410 : 194.5	194.9
E5	1410 : 195.5	194.9
C6	1350 : 411.5	411.4
A5, B6 and C7	1470 : 290.5	291.0

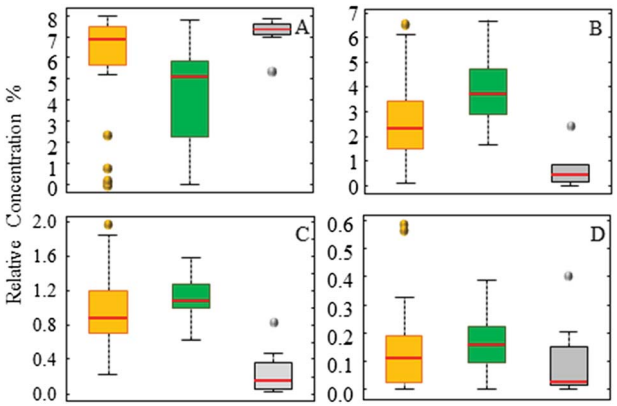

Fig. 4 Box-and-whisker plots showing the relative content of four chemical species in Taiwanese oolong tea (orange; $n = 46$), Taiwanese green tea (green; $n = 8$), and Taiwanese black tea (grey; $n = 10$). Potential chemomarkers: (A) 1350 s:194.5 u e^{-1}; (B) 1470.0 s:290.5 u e^{-1}. Selected buckets: (C) 1410 s:481.5 u e^{-1}; (D) 630.0 s:366.5 u e^{-1}. The % content was calculated in relation to the sum of the values of all the buckets selected during statistical analysis (the sum corresponds to 100%). In this analysis, the general bucketing method (i) was used.

catechin epimers (represented by vectors A5, B6, and C7). The group bucketing method also revealed a trend in the relative content of a substance corresponding to the peak at the m/z 174.9 u e^{-1}, which was putatively assigned to theanine. The highest intensity of this signal was observed in green tea samples.

Notably, although the selection of buckets according to the sequence shown in Fig. 1 was based on the division into type classes (oolong/green/black tea), the obtained PCA results show satisfactory separation of Taiwanese from non-Taiwanese teas (*cf.* Fig. 3D and E for black and green teas, respectively). This is of potential significance because expensive teas produced in Taiwan are sometimes adulterated by mixing with low-cost tea leaves of unknown origin.

3.3. Chemomarkers for classes

As is evident from the PCA plots (Fig. 3A–C) discussed above, the relative content of caffeine and catechin in tea is linked with the degree of fermentation. This is also clear from the comparison of oolong, green and black tea in box-and-whisker plots (Fig. 4). A high caffeine content was recorded for black and oolong teas. Moreover, the relative content of caffeine varied among the samples of green tea. Several samples of oolong tea had unusually low content of caffeine (oolong tea of fourth grade quality (#45 and #54), organic oolong tea (#42), and an uncharacteristic oolong tea brand (#104); *cf.* outliers in Fig. 4A). To verify the robustness of using caffeine as well as epi/catechin as chemomarkers for tea classes, we attempted using tests of statistical significance. Unfortunately, some of the datasets represented in Fig. 4 cannot be modelled by normal distribution – as verified using Jarque–Bera, Shapiro–Wilk and Doornik Chi–Square tests. Therefore, a (non-parametric) two-sample Kolmogorov–Smirnov test was used to compare the datasets. This analysis rejected the null hypothesis that the values of caffeine in the oolong/green tea samples as well as the black/green tea samples are drawn from groups with the same distributions ($p = 0.03$ and 0.01, respectively). The difference in the distribution of the caffeine content in black tea samples and oolong tea samples was not confirmed with this test ($p = 0.08$). Comparisons of catechin content distribution between the samples (oolong/green/black tea; *cf.* Fig. 4B) led to rejections of the null hypothesis, proving the reliability of this chemomarker. However, no significant differences were spotted with respect to a randomly selected bucket at 630.0 s:366.5 u e^{-1} (*cf.* Fig. 4D; $p > 0.05$). This further shows that the selection of buckets according to the lowest-RSD criterion (*cf.* Fig. 1), combined with PCA, may be useful in the search for chemomarkers.

3.4. Quantification of chemomarkers in selected samples

Fig. 5 shows overlaid total ion currents (TICs) of green, oolong and black tea obtained in the positive-ion mode. The retention times of the peaks in the TICs of real samples correspond to the peaks (m/z 195, 291, 307, 443 and 459 u e^{-1}) in the extracted ion currents (EICs) obtained for a mixture of chemical standards (Fig. S4†). The mass spectra corresponding to the eluting species revealed fragmentation patterns similar to the spectra obtained for the standard mixture, allowing for the identification of the species of interest. Following the identification of several unknown peaks, including those related to prominent chemomarkers, the LC-MS method was used in quantitative analysis of the corresponding compounds (Table S2†). Concentration values of the target species were determined based on duplicate chromatographic separations of the same extracts. In the case of epigallocatechin, detected in tea extracts, quantification was performed in relation to an isomeric compound gallocatechin. Table S3† presents the

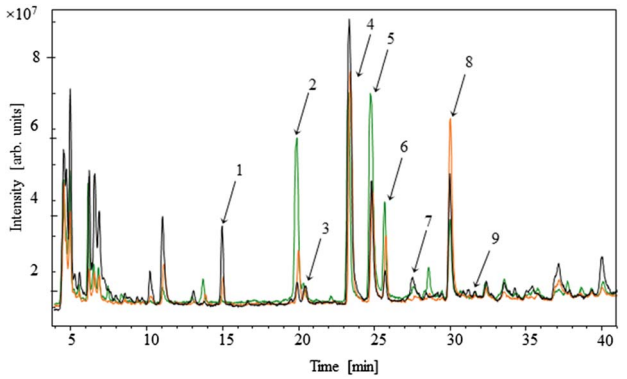

Fig. 5 Total ion chromatograms of green (green line, #1), oolong (orange line, #2) and black tea (black line, #15). Peaks: (1) gallocatechin; (2) epigallocatechin; (3/6) catechin or epicatechin; (4) caffeine; (5/7) epigallocatechin 3-gallate or gallocatechin 3-gallate; (8/9) epicatechin 3-gallate or catechin 3-gallate.

results of the measurements for three randomly selected representative samples of green, oolong and black tea. The highest content of catechin substances was observed in a green tea sample (50.2 mg g^{-1}) and of caffeine in a black tea sample (16.4 mg g^{-1}).

4. Conclusions

This work demonstrates the possibility to use LC-ESI-MS in conjunction with adequate chemometric treatment to characterize tea samples. Teas manufactured in Taiwan and other countries, as well as different kinds of Taiwanese tea (black, oolong, and green) could be differentiated based on their chemical signatures. A data reduction (according to the lowest-RSD criterion) was used in conjunction with PCA to facilitate the observation of such differences. The results demonstrate a dependence of metabolic profiles of tea on the manufacturing processes with an emphasis on the degree of fermentation. The authors think it would be useful to apply this method to profile other teas from various countries to ensure high quality, and counteract adulteration of the expensive brands. Profiling larger numbers of teas using the same method would enable setting cut-off values for different chemomarkers and tea brands; this could be then followed with blind tests, and transferring the methodology to the routine use. It is also suggested the proposed route of data treatment holds potential for selection of chemo- and bio-markers in chemotaxonomic and biomedical studies.

Acknowledgements

We thank the National Science Council of Taiwan, the National Chiao Tung University, and the Center for Interdisciplinary Science (NCTU) for the financial support of this work. We also thank Paraytec Ltd. (York, UK) for providing us the ActiPix D100 UV imaging detector.

Notes and references

1 A. Jain, C. Manghani, S. Kohli, D. Nigam and V. Rani, *Toxicol. Lett.*, 2013, **220**, 82–87.
2 B. Dhiman and M. Singh, *Planta Med.*, 2003, **69**, 882–884.
3 S. L. Liu, Y. S. Tsai and A. S. Ou, *J. Food Drug Anal.*, 2010, **18**, 34–43.
4 L. Mei, P. Lundin, M. Brydegaard, S. Gong, D. Tang, G. Somesfalean, S. He and S. Svanberg, *Appl. Opt.*, 2012, **51**, 803–811.
5 J.-Y. Cho, M. Mizutani, B. Shimizu, T. Kinoshita, M. Ogura, K. Tokoro, M.-L. Lin and K. Sakata, *Biosci., Biotechnol., Biochem.*, 2007, **71**, 1476–1486.
6 H.-F. Wang, Y.-S. Tsai, M.-L. Lin and A.-S. Ou, *Food Chem.*, 2006, **96**, 648–653.
7 J.-J. Wu, M.-T. Chiang, Y.-W. Chang, J.-Y. Chen, H.-T. Yang, C.-K. Lii, J.-H. Lin and H.-T. Yao, *J. Food Drug Anal.*, 2011, **19**, 289–300.
8 Y. Wang, Q. Li, Q. Wang, Y. Li, J. Ling, L. Liu, X. Chen and K. Bi, *J. Agric. Food Chem.*, 2012, **60**, 256–260.
9 K. Podwojski, A. Fritsch, D. C. Chamrad, W. Paul, B. Sitek, P. Mutzel, C. Stephan, H. E. Meyer, W. Urfer, K. Ickstadt and J. Rahnenfuhrer, *Bioinformatics*, 2009, **25**, 758–764.
10 R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, *BMC Genomics*, 2006, **7**, 142.
11 F. C. Marhuenda-Egea, R. D. Gonsálvez-Álvarez, B. Lledo-Bosch, J. Ten and R. Bernabeu, *Anal. Chem.*, 2013, **85**, 3053–3058.
12 K. R. Gabriel, *Biometrika*, 1971, **58**, 453–467.
13 F. W. Young, *Principal Components: BiPlots*, 1999, http://forrest.psych.unc.edu/research/vista-frames/help/lecturenotes/lecture13/biplot.html.

