

1**Title:**

2

3**Rapid and direct volatile compound profiling of black and green teas (*Camellia sinensis*) from**
4**different countries with PTR-ToF-MS**

5

6**Authors**

7Sine Yener^{1,2}, José A. Sánchez-López^{3,4}, Pablo M. Granitto⁵, Luca Cappellin¹, Tilmann D. Märk⁷,
8Ralf Zimmerman^{4,6}, Günther Bonn², Chahan Yeretizian³, Franco Biasioli¹

9

10**Affiliations**

11

12¹Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund
13Mach, San Michele all'Adige, Italy

14²Leopold Franzens University Innsbruck, Institute of Analytical Chemistry & Radiochemistry,
15Innsbruck, Austria

16³Zurich University of Applied Sciences (ZHAW), Institute of Chemistry and Biological Chemistry,
178820 Wädenswil, Switzerland

18⁴Joint Mass Spectrometry Centre, Chair of Analytical Chemistry, Institute of Chemistry, University
19of Rostock, D-18059 Rostock, Germany

20⁵CIFASIS, French Argentine International Center for Information and Systems Sciences, UNR–
21CONICET, Bv 27 de Febrero 210 Bis, 2000 Rosario, Argentina

22⁶Joint Mass Spectrometry Centre, Cooperation Group Comprehensive Molecular Analytics/CMA,
23Helmholtz Zentrum München, D-85764 Neuherberg, Germany

24⁷Leopold-Franzens University of Innsbruck, Institute of Ion Physics and Applied Physics,
25Innsbruck, Austria

26

27**Running title:** Discrimination of black and green teas with PTR-ToF-MS

28**Highlights**

- 29 • 101 teas, both leaves and infusions, of different origins have been analysed by PTR-ToF-MS
30 • Black and green teas were separated based on VOC emissions
31 • The volatile profiles of teas are highly affected by type and brewing process
32 • Multivariate data analyses indicate a possible classification according to geographical origin

33**Abstract**

1

2

34 Volatile profiles of 63 black and 38 green teas from different countries were analysed with Proton
35 Transfer Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS) both for tea leaves and tea
36 infusion. The headspace volatile fingerprints were collected and the tea classes and geographical
37 origins were tracked with pattern recognition techniques. The high mass resolution achieved by ToF
38 mass analyser provided determination of sum formula and tentative identifications of the mass
39 peaks. The results provided successful separation of the black and green teas based on their
40 headspace volatile emissions both from the dry tea leaves and their infusions. The volatile
41 fingerprints were then used to build different classification models for discrimination of black and
42 green teas according to their geographical origins. Two different cross validation methods were
43 applied and their effectiveness for origin discrimination was discussed. The classification models
44 showed a separation of black and green teas according to geographical origins the errors being
45 mostly between neighbouring countries.

46

47 **Key words:** tea aroma, tea leaf, tea infusion, volatile profiling, headspace volatile fingerprinting,
48 geographic origin classification

491 **Introduction**

50 In tea production, the leaves of the tea plant *Camellia sinensis* are used as the same starting material
51 but the differences in the processing techniques result in a wide range of characteristic teas with
52 distinct sensory properties. According to the way of processing, teas are usually classified into three
53 big groups based on their fermentation degrees: non-fermented (green and white), semi-fermented
54 (oolong) and fully fermented (black tea including pu-erh tea) [1]. There are several tea producing
55 countries in the world. The main five tea producing countries are China, India, Kenya, Sri Lanka
56 and Turkey [2]. Each country has different regions with their own climate and tea processing
57 methods which characterize color, appearance and flavor of the final product. For this reason, most

58tea products are marketed with the indication of the production region for product authentication
59and valorization.

60Aroma compounds play an important role for consumer preferences and perception of tea. Starting
61with the fresh tea leaves, which have a greenish and unripe odour, the characteristic tea aroma is
62developed during tea leaves processing. The most investigated volatile compounds (VOCs) in tea
63mainly consist of non-terpenoid and terpenoid components; the former are products of fatty acid
64degradation and provide the fresh green flavour, the latter are mostly monoterpene alcohols which
65give a floral sweet aroma [3, 4].

66Various studies have been conducted in the field of tea aroma research as recently reviewed by Yang
67et al. (2013) [5]. In short, gas chromatography-mass spectrometry (GC-MS) is generally used as a
68reference method in order to identify and quantify VOCs. The odour characteristics of volatiles
69have been detected with aroma dilution and GC-olfactometry; and recently electronic nose
70techniques have been used for fast analysis of tea aroma. These methods have allowed analysing the
71volatile profiles of teas at different fermentation degrees and also to classify green, black and
72oolong teas according to their geographical origins [6-11]. Among them, GC-MS has turned out to
73be the most accurate and effective method for identification, separation and quantification of
74volatile compounds; however it requires capturing volatiles by various extraction methods which
75are generally time consuming and their efficiency depends on the characteristics and limitations of
76the analytical approach (e.g. the absorption and desorption of volatiles from a specific material in
77the case of SPME)[12].

78To link sensory perception of tea with instrumental data, direct and non-destructive instrumental
79analysis of volatiles can be considered to be the most appropriate approach because it provides a
80direct estimation of the VOCs released of from tea and that reach the human olfactory system. In
81this regard, proton transfer reaction-mass spectrometry, PTR-MS, provides an efficient approach as
82a direct injection, soft chemical ionization method for the analysis of VOCs at trace levels. The
83direct injection method requires no sample pre-treatment which allows real-time monitoring of

84VOCs [13, 14] without making any changes in the volatile composition of samples. The technique
85uses H_3O^+ ions for protonation of VOCs with proton affinities **higher** than that of water which can
86be further analysed by a quadruple or a time-of-flight (ToF) mass analyser [15]. ToF mass analysers
87provide high sensitivity that leads to detection of volatiles at ppt levels and high mass resolution
88which allows, in most cases, the identification of the sum formula of the observed peaks [13].

89PTR-MS allows collecting the overall mass spectral fingerprints of the samples which can be
90further processed with advanced data analysis tools for successful discrimination and classification
91of the food products [16]. To the best of our knowledge, neither a study has been conducted on the
92analysis of volatile compounds emitted from various tea types by PTR-MS nor was this method
93applied for discrimination of teas from different geographical origins.

94With this study, we aim to apply PTR-ToF-MS, for the first time, for aroma profiling of black and
95green tea samples, both leaves and brew, from different countries and to investigate the possibility
96of origin tracing on the basis of their geographical origins with the aid of chemometric tools.

972 **Materials and Methods**

982.1 **Tea samples**

99In total, 101 commercially available pure tea samples, without addition of flavouring agents, from
10016 different countries (Table 1) were purchased from the market; 63 black teas and 38 green teas.
101The samples were stored in their original bags at room temperature before analysis. Trademarks and
102producers were kept confidential but the commercial names, origins and other characteristics of the
103tea samples are provided in Supplementary material S1.

1042.2 **Analysis of tea volatiles by PTR-ToF-MS**

105The volatile compounds of dry tea leaves and their infusions were analysed by PTR-ToF-MS by
106direct injection headspace analysis without destructing the original samples. For the analysis of dry
107tea leaves, 500 mg tea leaves were weighted into 22-ml glass vials (Supelco, Bellefonte, PA) and 3

108 replicates were prepared for each tea sample. Tea brewing was performed by applying a 3 min fixed
109 infusion time for all tea samples. Deionized hot water (25 ml, 85°C) was used for brewing of tea
110 leaves (400 mg) in 40 ml amber vials (Supelco, Bellefonte, PA). The liquid infusion was taken right
111 after brewing by a micropipette and 2 ml of aliquots were transferred into 22-ml glass vials. Each
112 tea sample was brewed 3 times and each brew was analysed in duplicate.

113 The headspace measurements were performed by using a commercial PTR-ToF-MS 8000
114 instrument (Ionicon Analytik GmbH, Innsbruck, Austria). The instrumental conditions in the drift
115 tube were set as following: drift voltage 550 V, drift temperature 110°C, drift pressure 2.33 mbar
116 affording an E/N value (electric field strength/gas number density) of 140 Townsend ($1 \text{ Td} = 10^{17} \text{ V.cm}^{-2}$). All the vials containing samples and blank vials (air for tea leaves and hot water for tea
118 brews) were incubated at 37°C for 30 min before headspace analysis. The headspace mixture was
119 directly injected into PTR-MS drift tube with a flow rate of 40 sscm via a PEEK tube at 110°C.
120 Sample injection was performed with a multipurpose autosampler (Gerstel GmbH, Mulheim am
121 Ruhr, Germany). A different sample was analysed every 5 min. Each sample was measured for 30 s,
122 at an acquisition rate of one spectrum per second. The measurement order was randomized while
123 measuring the volatile emissions of tea leaves and tea brews.

124 2.3 Data processing and analysis

125 2.3.1 Treatment of mass spectrometric data

126 Data processing of ToF spectra included dead time correction, internal calibration and peak
127 extraction steps performed according to a procedure described elsewhere [17] to reach a mass
128 accuracy ($\geq 0.001 \text{ Th}$) which is sufficient for sum formula determination. The baseline of the mass
129 spectra was removed after averaging the whole measurement and peak detection and peak area
130 extraction was performed by using modified Gaussian to fit the data [18]. Whenever a peak was
131 detected, the volatile concentrations were calculated directly via the amount of detected ions in
132 ppbv (part per billion by volume) levels according to the formulas described by Lindinger et al.

133(1998) [13] by assuming a constant reaction rate coefficient ($k_R=2\times 10^{-9}$ cm³/s). For H₃O⁺ as a
134primary ion, this introduces a systematic error for the absolute concentration for each compound
135that is in most cases below 30% and can be accounted for if the actual rate constant is available
136[19].

1372.3.2 Selection of mass peaks

138The direct injection headspace analysis of tea (leaves and infusion) samples resulted in identifying
139455 mass peaks in the range 15-300 *m/z*. After eliminating the interfering ions (O₂⁺, NO⁺ and water
140clusters) and their isotopologues, 438 mass peaks remained for further analysis. The signals
141belonging to blank vials were subtracted from the whole data set (air from tea leaf emissions, water
142from infusion emissions). A concentration threshold of 0.1 ppb was set for further reduction of noise
143in the data matrices. After this step 257 mass peaks x 303 (i.e. 101 samples, three biological
144replicates) data points were left to build the matrix containing tea leaf emissions; 162 mass peaks x
145606 (i.e. 101 samples, three infusions, two analytical replicates) data points were left for tea
146infusion data matrix. These final data matrices were used for univariate and multivariate data
147analysis methods.

148After mass peak selection and extraction, tentative peak identification was performed by using an
149in-house library developed by the authors where the peak annotations were done automatically with
150the scripts developed under R programming language [20].

1512.3.3 Statistical analyses

152The significant differences between tea types were calculated using ANOVA (99% confidence
153level) and the pairwise comparison was performed with Tukey's test to highlight these differences
154with letter annotations.

155As a first step, the final data matrices were subjected to principal component analysis (PCA).
156Secondly, Random Forests (RF), Penalized Discriminant Analysis (PDA), Support Vector Machines

157(SVM) and Discriminant Partial Least Squares (dPLS) classification methods were applied for
158sample discrimination [21] and their classification power was compared.

159Two types of validation methods were tested for each classification method: a simple 6-fold cross
160validation and Leave-Group-Out (LGO) cross-validation. The six-fold cross-validation was
161performed by randomly dividing the whole data set into 6 folds. One of the folds was removed at
162each time and used as a test set where the rest of the data (the train set) was used to build the
163discriminant method and predict the origins of samples. Using this cross validation method,
164analytical or biological replicates of the same tea sample can be at the same time in both the train
165and test sets. With the highly flexible classification methods used in this work, this can easily leads
166to overfitting the data and to produce biased estimates of classification errors. This effect was
167verified in preliminary experiments (**not shown**) and the method was discarded. In the case of the
168more elaborated LGO cross-validation, the analytical and biological replicates of each tea sample
169were considered as a group when discriminating tea types and geographical origin. Each time, one
170group was removed from the full dataset and used as a test set. Mean classification errors and
171confusion matrices were used to evaluate the performance of each classification method. All the
172multivariate data analyses were performed by using the scripts and packages developed under R
173programming language [20].

1743 **Results and Discussion**

1753.1 **Volatile profiling of black and green teas and discrimination based on tea type**

176One-way ANOVA of the mass peaks extracted in black and green tea headspace, showed 135 mass
177peaks significantly different ($p < 0.01$ with Bonferroni correction) between emissions of black and
178green tea leaves and 125 mass peaks between their infusions. Among the mass peaks extracted, 62
179of them were tentatively identified as one or more volatile compounds based on their presence in

180dry tea leaves and brews reported in literature. The details of the tentatively identified mass peaks
181are shown in Table 2 with their average concentrations in black and green tea leaves and infusions.

182The leaves of different tea types showed greater volatile emissions as compared with infusions.

183Various terpenes and their fragments dominated the volatile emission of tea leaves, followed by
184esters/acids and aldehydes/ketones. In particular, green tea leaves emitted more terpenes and
185sulphur compounds than black teas. The most abundant volatile compounds in the headspace of
186green tea infusions were sulphur compounds, aldehydes/ketones and terpenes. The headspace of
187black tea infusions contained aldehydes/ketones the highest; sulphur compounds, terpenes and
188alcohols were other most abundant chemical groups.

189Some distinct differences and similarities can be pointed out between black and green teas: the most
190abundant sulphur compound detected in both tea infusions was tentatively identified as dimethyl
191sulphide. It has been reported that this sulphur compound improves the flavour of green teas
192harvested in spring [3]. The information about the season when the green teas were picked was not
193available for all the tea samples but for some of the black teas. Interestingly, we observed that the
194black teas that had the highest dimethyl sulphide contents were indeed picked during spring (e.g.
195sample no 102, 110, and 116 in Supplementary file 1).

196We observed that the percentage of total monoterpenes and their fragments in the headspace of
197black tea infusions (~20%) was higher than the amount emitted from green tea infusion (~12%).
198Terpenes, especially monoterpenes, are responsible for the characteristic floral odour of tea [22].
199Important aroma compounds derived from breakdown of carotenoids during black tea processing
200like linalool, geraniol, linalool oxide and ionone [3] were also higher in the headspace of black teas
201and their infusions than in green teas. Most of the monoterpenes and derived compounds were
202significantly lost during tea brewing; in particular linalool oxide (m/z 171.133) in green tea
203infusions.

204Vanillin was previously reported to be one of the compounds of highest flavour dilution factor (FD)
205in black tea infusion [23]. In our study, the peak corresponding to vanillin was negligible in green
206tea infusions, but clearly observable in black teas with little effect of brewing.

207When PCA was performed, the first three principal components provided a good separation of black
208teas from green teas based on the volatile emissions both from dry leaves and infusions (Figure 1a,
209b). The first PCs explain 53.2 and 54.7% variances for the dry tea leaves and infusions, respectively.
210This reflects the high variance between black and green tea volatile emissions as well as within each
211tea type (black or green) depending on the different production methods and origins. The release
212mechanisms of volatiles might be influenced by matrix characteristics (*i.e.* leaf shape and size) as
213teas can be produced in various shapes. For example, the green teas can be shaped like needle,
214twisted, flat, round, compressed shape or even as ground powder as a results of fixing and drying
215methods. Besides, leaf disruption also occurs in cutting and rolling steps of black tea production
216that leads to grading of black teas according to leaf size [22].

217Figure 2a and 2b show score plots of the first two PCs of tea leaves and tea infusions (loadings of
218the first two components of tea leaves and infusions are provided in Supplementary file S2).
219According to these Figure 2a and 2b, some black tea samples with broken leaves (sample numbers
2201, 10, 30, 43, 112, 146-148) were closely located and separated from others. These samples were
221characterized by the mass peaks at m/z 59.049, 85.065, 97.065, 99.081, 111.081, 113.096, 115.074,
222115.112, 139.113, 141.127 and 143.144 which were mostly attributed to aldehydes and ketones;
223mass peaks at m/z 101.096 and 87.080 to alcohols and mass peak at m/z 169.126 to geranic acid in
224the headspace of dry tea leaves. In addition, mass peaks; m/z 71.049 (butenal), 77.058 (propandiol),
225129.099 (hexenyl formate), 127.112 (methylheptenone) had high loadings in the headspace of tea
226infusions with broken leaf shape. Broken and smashed tea may release more catechins than firmly
227pressed tea leaves and they may undergo heavier oxidation [24]. Broken leaves also provide a larger
228surface area during fermentation favouring enzymatic (*i.e.* glycosidases, fatty acid hydroperoxide

229lyase) activity for production of volatile aldehydes [25]. These findings indicate the importance of
230leaf shape on volatile compound generation and their extraction during the infusion process.

231PCA showed a **relatively** good separation of black and green teas by using three principal
232components (Figures 1a-b, 2a-b). To be able to assess the performance of discrimination, we applied
2334 classification models by using two different cross validation methods **for discrimination of black**
234**and green teas. According to LGO cross validation the average errors for classification black and**
235**green tea leaves were 2.6, 3.9, 1.3 and 3.6 %; the average errors for classification black and green**
236**tea infusions were 0.6, 0.2, 0.0 and 0.5 % obtained by RF, PDA, SVM and dPLS classification**
237**models, respectively. In general, all the classification techniques showed very good classification**
238**efficiency with an average error rate less than 4.0% for differentiating black and green tea volatile**
239**profiles emitted from leaves and infusions. In all cases, the classification errors were lower for tea**
240**infusions than tea leaves.**

2413.2 **Geographical origin discrimination with supervised classification methods**

242The results described above highlighted significant differences between black and green tea aroma
243profiles and successful separation of large number of tea samples according to tea type. However it
244would more relevant to demonstrate that the volatile composition of tea might be related to its
245geographical origin, as well. For this reason, we applied supervised classification methods on the
246black and green tea volatile profiles in order to differentiate them according to their origins.

247To get a more representative data set for classification studies, we selected origins (countries)
248represented by at least 4 different teas. Black teas from China, India, Sri Lanka and Nepal (50
249samples) were included for classification of black teas; China, India, Japan and Korea (32 samples)
250were selected for classification of green teas. Each classification algorithm ended up with an
251average classification error and a confusion matrix where the original tea origins were compared
252with the origins assigned by the classification method. The classification methods were applied on

253normalized volatile concentrations with LGO cross validation tests. **The normalized concentrations**
254**were obtained by normalizing each mass spectrum to unit area as described in [26].**

255The classification performances obtained by using emissions of the tea leaves and tea infusions
256were similar and they provided relatively good separations which were between 30-50%. Due to the
257fact that, the tea infusions are the final consumed products, in the following discussion, we will
258focus on the classification models tested for black and green tea infusions.

259**Table 3** shows confusion matrices and the performances of the classification models applied on
260black and green tea infusions. Among the 50 black teas from 4 countries tested, the lowest
261prediction error was around 32% obtained by RF; the same method also provided the highest
262classification performance for green tea infusions prepared by 32 samples from 4 countries. Among
263the black teas, teas from Sri Lanka were classified with lowest errors followed by India, China and
264Nepal. In the case of green teas, Chinese teas had the lowest error followed by Japan, India and
265Korea. The confused tea samples were mostly from the neighbouring countries. For instance,
266Korean green teas were confused with Chinese and Japanese green teas but not with Indian green
267teas with RF method. This finding is not surprising because political borders are not likely to affect
268tea quality while climate, growing conditions, picking method and processing traditions [3, 22, 27]
269are the key factors for differentiating tea classes and their characteristics. Unfortunately we were not
270able to find better geographical indications for many samples.

271Similar cases have been reported in literature with various classification performances when
272different tea samples were discriminated according to geographical origin based on their volatile
273profiles. Togari, Kobayashi and Aishima (1995) [10] performed the first study on the geographical
274origin determination of different tea categories based on their volatile profiles. The study included
275GC-MS analysis of 44 tea samples where tea volatiles were extracted by simultaneous dilution and
276extraction (SDE) method by mixing the tea samples with hot water. Black teas from India (8), Sri
277Lanka (4) and Japan (1) were successfully classified by supervised pattern recognition techniques,
278however neither oolong (China (10) and Taiwan (4)) nor green teas (15, from different regions of

279Japan) could be classified according to origin. Kovács et al., (2010) [8] applied electronic nose
280technology with electronic tongue and sensory assessment for geographical origin discrimination of
281five Sri Lankan teas. When electronic tongue responses of tea infusion headspace was treated with
282linear discriminant analysis, 100% correct classification was obtained for middle and low elevation
283regions (n=3) however two samples from high elevation showed overlapping. Ye, Zhang and Gu
284(2011) [11] analysed volatile profiles of 23 green tea samples produced in two different regions of
285China with SPME/GC-MS via extracting the volatiles from tea powder. They could classify the
286production areas of tea samples. Lee et al. (2013) [9] analysed 24 green tea samples from 8 different
287countries (China (7), India (1), Japan (6), Kenya (2), Korea (4), Sri Lanka (2), Tanzania (1) and
288Vietnam (1)) with GC-SPME method nevertheless no relationship has been found between country
289of origin and aroma where specific information about the samples other than origin was not known
290for the tested tea samples. In another study, 38 tea samples from China (2 oolong, 2 green, 3 black),
291Japan (5 green, 3 black, 2 oolong), India (5 black), Sri-Lanka (5 black), and Chinese Taipei (6
292oolong, 2 black) were analyzed by GC-MS and they were classified according to their origins by
293clustering methods [6]. Lastly, four varieties of oolong teas were analysed by olfaction and
294gustation sensing systems, the samples were classified according to producing regions by using the
295information each sensing system provided. When all information was merged with data fusion
296techniques, the discrimination power increased compared to individual classification performances
297suggesting the possibility to use these systems with multivariate methods for discriminating and
298classifying tea samples [7].

299When our results and the literature were considered together, different tea types from various
300countries can be discriminated to some extent according to geographical origin based on their
301volatile emissions from dry tea leaves or tea infusions. Moreover fermented tea products are better
302classified than non-fermented and semi-fermented teas which was also observed from our results
303when we compare the classification efficiencies of black and green teas.

304When black tea infusions from India (Assam (9) and Darjeeling (12)), China (Anhui (3) and Yunnan
305(7)), Sri Lanka (all country) and Nepal (all country) were classified according to tea producing
306regions a significant improvement on the classification performance was observed providing 15%
307average error rate (confusion matrices not shown). The results indicated 4 classes: China Anhui
308(class 1); Sri Lanka and India Assam (class 2) Nepal and Darjeeling (class 3) and China Yunnan
309(class 4) by showing the geographically close regions in the same group.

310Overall, these findings point out that the regions might be better differentiators instead of the
311country and the regions closely located to each other share more similar properties and they are
312likely to create a group. Besides, there might be other factors affecting the volatile composition of
313different types of tea in addition to geographical location such as the age of the tea plant, plucking
314(fine or coarse), plucking season, tea processing, packaging of tea, conditions during storage and
315storage time, which should be taken into consideration.

3164 **Conclusions**

317In this study, for the first time, the volatile profiles of black and green teas from 12 different
318geographical origins were analysed by PTR-ToF-MS. The volatile compounds of a large sample set
319(101 samples, with replicates, both leaves and infusions) were analysed by direct injection of the
320headspace without altering the original tea components and destructing the original sample. The
321high mass resolution and sensitivity achieved by the mass analyser enabled annotation of sum
322formulas to the detected mass peaks. Tentative identifications lead defining important aroma
323compounds in black and green tea volatile emissions and pointed out the differences among them.

324Black and green teas were correctly classified by the volatile compounds emitted from tea leaves
325and their infusions independent from their geographical origins. Classification models were built to
326predict the geographical origins of black and green teas. Results provided a good separation of tea
327origins; however countries geographically close to each other were most likely to be confused.
328Preliminary analysis indicated that a better discrimination of tea samples might have been achieved

329if teas were classified according to production region rather than just country of origin. This was not
330feasible here, since information about production region was available only for a limited number of
331samples.

332Our results showed that PTR-ToF-MS fingerprints combined with multivariate statistical techniques
333provided successful evaluation of tea products. Considering the very promising results obtained so
334far, in discriminating for processing and country, it seems highly warranted to collect significantly
335more detailed information about the individual tea samples, for future studies. This may include e.g.
336information on production region, producer, harvesting season, post-harvest treatment and age of
337the product. It may also be significant to investigate the effect of tea leave shape and infusion
338conditions. Finally, it is also important to direct our interest towards the consumer, by analysing the
339volatile compounds release from the nospace and analysed by PTR-ToF-MS, when a tea product
340is being consumed, and conducting sensory profiling as well. Combining such a large spectrum of
341different data sets might currently seem to be a veritably challenging task; we believe this will need
342to be approached in steps towards a more complete understanding of the factors affecting tea aroma
343profiles.

344

345**Acknowledgements**

346This research has been funded by PIMMS (Proton Ionisation Molecular Mass Spectrometry) ITN,
347which is supported by the European Commission's 7th Framework Programme under Grant
348Agreement Number 287382.

349

350**References**

- 351[1] L. Zhang, Z. Zeng, C. Zhao, H. Kong, X. Lu, G. Xu, A comparative study of volatile
352 components in green, oolong and black teas by using comprehensive two-dimensional gas
353 chromatography-time-of-flight mass spectrometry and multivariate data analysis, *Journal of*
354 *chromatography. A*, 1313 (2013) 245-252.
355[2] FAOSTAT, available from http://faostat3.fao.org/browse/Q/*/E, accessed on 02.11.2015.
356[3] Z.M.. Chen, H.F. Wang, X.Q. You, N. Xu, in: Y.S. Zhen (Ed.) *Tea Bioactivity and Therapeutic*
357 *Potential*, Taylor & Francis, New York, USA, 2005, pp. 89-120.

- 358[4] P. Pripdeevech, S. Wongpornchai, in: V.R. Preedy (Ed.) *Tea in Health and Disease Prevention*,
359 Elsevier, USA, 2013, pp. 307-322.
- 360[5] Z. Yang, S. Baldermann, N. Watanabe, Recent studies of the volatile compounds in tea, *Food*
361 *Research International*, 53 (2013) 585-599.
- 362[6] S. Baldermann, Z. Yang, T. Katsuno, V.A. Tu, N. Mase, Y. Nakamura, N. Watanabe,
363 Discrimination of Green, Oolong, and Black Teas by GC-MS Analysis of Characteristic
364 Volatile Flavor Compounds, *American Journal of Analytical Chemistry*, 05 (2014) 620-632.
- 365[7] Q. Chen, C. Sun, Q. Ouyang, Y. Wang, A. Liu, H. Li, J. Zhao, Classification of different
366 varieties of Oolong tea using novel artificial sensing tools and data fusion, *LWT - Food*
367 *Science and Technology*, 60 (2015) 781-787.
- 368[8] Z. Kovács, I. Dalmadi, L. Lukács, L. Sipos, K. Szántai-Kőhegyi, Z. Kókai, A. Fekete,
369 Geographical origin identification of pure Sri Lanka tea infusions with electronic nose,
370 electronic tongue and sensory profile analysis, *Journal of Chemometrics*, 24 (2010) 121-130.
- 371[9] J. Lee, D.H. Chambers, E. Chambers, K. Adhikari, Y. Yoon, Volatile aroma compounds in
372 various brewed green teas, *Molecules*, 18 (2013) 10024-10041.
- 373[10] N. Togari, A. Kobayashi, T. Aishima, Pattern recognition applied to gas chromatographic
374 profiles of volatile components in three tea categories, *Food Research International*, 28 (1995)
375 495-502.
- 376[11] N. Ye, L. Zhang, X. Gu, Discrimination of Green Teas from Different Geographical Origins by
377 Using HS-SPME/GC-MS and Pattern Recognition Methods, *Food Analytical Methods*, 5
378 (2011) 856-860.
- 379[12] G. Reineccius, *Flavor Analysis*, in: G. Reineccius (Ed.) *Flavor Chemistry and Technology*,
380 Second Edition, CRC Press, FL, USA, 2005, pp. 33-72.
- 381[13] W. Lindinger, A. Hansel, A. Jordan, On-line monitoring of volatile organic compounds at pptv
382 levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications,
383 food control and environmental research, *International Journal of Mass Spectrometry and Ion*
384 *Processes*, 173 (1998) 191-241.
- 385[14] R.S. Blake, P.S. Monks, A.M. Ellis, Proton-Transfer Reaction Mass Spectrometry, *Chemical*
386 *Reviews*, 109 (2009) 861-896.
- 387[15] R.S. Blake, C. Whyte, C.O. Hughes, A.M. Ellis, P.S. Monks, Demonstration of Proton-Transfer
388 Reaction Time-of-Flight Mass Spectrometry for Real-Time Analysis of Trace Volatile Organic
389 Compounds, *Analytical Chemistry*, 76 (2004) 3841-3845.
- 390[16] F. Biasioli, F. Gasperi, C. Yeretziyan, T.D. Märk, PTR-MS monitoring of VOCs and BVOCs in
391 food science and technology, *TrAC Trends in Analytical Chemistry*, 30 (2011) 968-977.
- 392[17] L. Cappellin, F. Biasioli, A. Fabris, E. Schuhfried, C. Soukoulis, T.D. Mark, F. Gasperi,
393 Improved mass accuracy in PTR-TOF-MS: Another step towards better compound
394 identification in PTR-MS, *Int. J. Mass Spectrom.*, 290 (2010) 60-63.
- 395[18] L. Cappellin, F. Biasioli, P.M. Granitto, E. Schuhfried, C. Soukoulis, F. Costa, T.D. Mark, F.
396 Gasperi, On data analysis in PTR-TOF-MS: From raw spectra to data mining, *Sens. Actuator*
397 *B-Chem.*, 155 (2011) 183-190.
- 398[19] L. Cappellin, T. Karl, M. Probst, O. Ismailova, P.M. Winkler, C. Soukoulis, E. Aprea, T.D.
399 Märk, F. Gasperi, F. Biasioli, On Quantitative Determination of Volatile Organic Compound
400 Concentrations Using Proton Transfer Reaction Time-of-Flight Mass Spectrometry,
401 *Environmental Science & Technology*, 46 (2012) 2283-2290.
- 402[20] R.Core Team, in: *R Foundation for Statistical Computing*, 2014.
- 403[21] F. Biasioli, P.M. Granitto, E. Aprea, D. Mott, C. Furlanello, T.D. Mark, F. Gasperi, Rapid and
404 non-destructive identification of strawberry cultivars by direct PTR-MS headspace analysis and
405 data mining techniques, *Sens. Actuators B, Chem.*, 121 (2007) 379-385.
- 406[22] Z.M. Chen, N. Xu, in: Y.S. Zhen (Ed.) *Tea Bioactivity and Therapeutic Potential*, Taylor &
407 Francis, New York, USA, 2005, pp. 35-56.

- 408[23] C. Schuh, P. Schieberle, Characterization of the Key Aroma Compounds in the Beverage
 409 Prepared from Darjeeling Black Tea: Quantitative Differences between Tea Leaves and
 410 Infusion, *J. Agric. Food Chem.*, 54 (2006) 916-924.
- 411[24] L.F. Wang, J.Y. Lee, J.O. Chung, J.H. Baik, S. So, S.K. Park, Discrimination of teas with
 412 different degrees of fermentation by SPME–GC analysis of the characteristic volatile flavour
 413 compounds, *Food Chemistry*, 109 (2008) 196-206.
- 414[25] M.E. Harbowy, D.A. Balentine, *Tea Chemistry, Critical Reviews in Plant Sciences*, 16 (1997)
 415 415-480.
- 416[26] P.M. Granitto, F. Biasioli, E. Aprea, D. Mott, C. Furlanello, T.D. Maerk, F. Gasperi, Rapid and
 417 non-destructive identification of strawberry cultivars by direct PTR-MS headspace analysis and
 418 data mining techniques, *Sensors and Actuators B*, 121 (2007) 379-385.
- 419[27] M. Johns, The evaluation of volatile quality factors in black tea : a report for the Rural
 420 Industries Research and Development Corporation, in: RIRDC publication, 1440-6845; no.
 421 98/46, Rural Industries Research and Development Corporation, Barton, A.C.T, 1998, pp. 1-58.
- 422[28] T. Yamanishi, The Aroma of Various Teas, in: G. Charalambous, G.E. Inglett (Eds.) *Flavor of*
 423 *Foods and Beverages: Chemistry and Technology Academic Press, New York, USA, 1978, pp.*
 424 *305-328.*
- 425[29] C. Alasalvar, B. Topal, A. Serpen, B. Bahar, E. Pelvan, V. Gokmen, Flavor characteristics of
 426 seven grades of black tea produced in Turkey, *J Agric Food Chem*, 60 (2012) 6323-6332.
- 427[30] K. Kumazawa, H. Masuda, Identification of Potent Odorants in Japanese Green Tea (Sen-cha),
 428 *J. Agric. Food Chem.*, 47 (1999) 5169-5172.
- 429[31] M. Kawakami, T. Shibamoto, The Volatile Constituents of Piled Tea: Toyama Kurocha, *Agric.*
 430 *Bioi. Chem.*, 55 (1991) 1839-1847.
- 431[32] W. Renold, R. Näf-Müller, U. Keller, B. Wilhalm, G. Ohloff, An Investigation of the Tea
 432 Aroma Part I. New Volatile Black Tea Constituents, *Helvetica Chimica Acta*, 57 (1974) 1301-
 433 1308.
- 434[33] P. Pripdeevech, T. Machan, Fingerprint of volatile flavour constituents and antioxidant
 435 activities of teas from Thailand, *Food Chemistry*, 125 (2011) 797-802.
- 436[34] T. Yamanishi, M. Kosuge, Y. Tokitomo, R. Maeda, Flavor Constituents of Pouchong Tea and a
 437 Comparison of the Aroma Pattern with Jasmine Tea, *Agric. Biol. Chem.*, 44 (1980) 2139-2142.
- 438[35] E.S. Kim, Y.R. Liang, J. Jin, Q.F. Sun, J.L. Lu, Y.Y. Du, C. Lin, Impact of heating on chemical
 439 compositions of green tea liquor, *Food Chemistry*, 103 (2007) 1263-1267.
- 440[36] P. Schieberle, C. Schuh, Aroma compounds in black tea powders of different origins-changes
 441 induced by preparation of the infusion, *Flavour Science: Recent Advances and Trends*, (2006)
 442 151-156.
- 443[37] H. Chi-Tang, Z. Xi, L. Shiming, Tea aroma formation, *Food Science and Human Wellness*
 444 4(2015) 9-27.
- 445

446 Figures

447 **Figure 1.** 3D PCA score plots of black and green tea leaves (a) and tea infusions (b). Black and
 448 green colours represent black and green teas, respectively.

449 **Figure 2.** 2D PCA score plots of black and green tea leaves and infusions. Black and green colours
 450 represent black and green teas, respectively. Due to the good repeatability of the analytical
 451 replicates, PCA was built via averaging the replicates. This improved the visualization of each

452sample. The numbers on the points indicate the sample codes given in Supplementary file S1. **Table 4531.** Distribution of tea samples according to tea types and countries of origin

Country (code)	Tea Types	
	<i>Black</i>	<i>Green</i>
Argentina (ARG)	1	-
China (CHI)	13	15
India (IND)	25	4
Indonesia (INDO)	3	-
Japan (JAP)	-	9
Kenya (KEN)	2	-
Korea (KOR)	1	4
Nepal (NEP)	4	1
Rwanda (RWA)	-	1
Sri Lanka (SRI)	8	1
Tanzania (TNZ)	1	1
Turkey (TUR)	3	-
Vietnam (VIE)	1	2
Zimbabwe (ZIM)	1	-
(Total)	63	38

454

455

456

457

458

459

460

461

462

463

33

34

464

465

466 **Table 2.** The average concentrations (ppb) of tentatively identified mass peaks in the headspace of black and green tea leaves and infusions. Peaks were selected on the basis of one-way ANOVA and the relative *p*-values are listed in the right columns.

Measured mass	The. mass	Sum formula	Chemical class	Tentative identification	Average concentration ± standart deviation (ppbv)		<i>p</i> -value	Average concentration ± standart deviation (ppbv)		<i>p</i> -value	Reference
					Black tea leaves	Green tea leaves		Black tea infusion	Green tea infusion		
33.0336	33.033	CH ₄ OH ⁺	Alcohols	Methanol	11756±3992	9555±3826	<0.001	170±82	120±63	<0.001	[28]
45.0333	45.034	C ₂ H ₄ OH ⁺	Aldehydes	Acetaldehyde	804±608	265±260	<0.001	416±158	108±58	<0.001	[28]
47.0491	47.049	C ₂ H ₅ OH ⁺	Alcohols	Ethanol	138±152	110±360	0.345	6±14	11±18	<0.001	[28]
49.0110	49.011	CH ₄ SH ⁺	Sulphur compounds	Methanetioli	0.9±0.5	0.5±0.3	<0.001	0.4±0.5	1.2±1.5	<0.001	[28]
59.0488	59.049	C ₃ H ₆ OH ⁺	Aldehydes/ketones	Propanal/acetone	340±479	265±784	0.293	115±99	92±85	0.002	[28]
61.0280	61.028	C ₂ H ₄ O ₂ H ⁺	Esters and acids	Acetic acid	804±605	390±848	<0.001	2.3±4.8	6.34±17	<0.001	[29]
63.0260	63.026	C ₂ H ₆ SH ⁺	Sulphur compounds	Dimethylsulfide	10±8	12±12	0.191	264±289	275±299	0.675	[28]
69.0333	69.034	C ₄ H ₄ OH ⁺	Furans	Furan fragment	7.0±7.4	4.7±6.1	0.004	7.2±7.8	2.0±3.0	<0.001	[30]
69.0697	69.070	C ₅ H ₈ H ⁺	Terpene fragment	Isoprene	183±145	220±217	0.067	145±76	35±21	<0.001	n.a.
71.0489	71.049	C ₄ H ₆ OH ⁺	Aldehydes	Butenal	19±17	12±8	<0.001	4.6±4.0	2.2±2.3	<0.001	[31]
73.0646	73.065	C ₄ H ₈ OH ⁺	Aldehydes	Methylpropanal	206±143	195±565	0.811	292±159	43±26	<0.001	[29]
75.0438	75.044	C ₃ H ₆ O ₂ H ⁺	Esters and acids	Propionic acid	96±103	42±42	<0.001	9.7±6.1	4.5±2.6	<0.001	[28]
79.0536	79.054	C ₆ H ₆ H ⁺	Aromatic hydrocarbons	Benzene	40±18	26±22	<0.001	5.3±7.3	4.1±7.7	0.043	[29]
81.0697	81.070	C ₆ H ₈ H ⁺	Terpene fragment	Cyclohexadiene (Terpene fragment)	612±905	423±461	0.035	29±24	9±12	<0.001	[28]
83.0854	83.086	C ₆ H ₁₀ H ⁺	Terpene fragment	Cyclohexene (Terpene fragment)	175±166	60±70	<0.001	71±63	19±23	<0.001	[28]
85.0646	85.065	C ₅ H ₈ OH ⁺	Aldehydes/Ketones	Pental/ pentenone	32±33	37±37	0.229	12±9	7.5±7.6	<0.001	[24, 29, 32]
87.0431	87.044	C ₄ H ₆ O ₂ H ⁺	Ketones	Butanedione	19±16	6.4±4.6	<0.001	2.7±3.1	1.3±0.7	<0.001	[29]
87.0802	87.080	C ₅ H ₁₀ OH ⁺	Alcohols	Pentenol	55±37	62±157	0.549	165±89	23±14	<0.001	[24, 29]
91.0559	91.058	C ₄ H ₁₀ SH ⁺	Sulphur compounds	Diethylsulphide/butanethiol (fragment)	7.9±3.6	4.4±2.7	<0.001	0.9±0.9	0.3±0.5	<0.001	[33]
93.0365	93.037	C ₃ H ₈ OSH ⁺	Sulphur compounds	Methylsulfanylethanol	8.6±3.7	5.6±2.6	<0.001	1.7±1.2	0.4±0.6	<0.001	n.a.
93.0698	93.070	C ₇ H ₈ H ⁺	Aromatic hydrocarbons	Toluene	55±88	34±34	0.014	5.9±7.5	2.2±2.3	<0.001	[28, 29, 32]
95.0173	95.016	C ₂ H ₆ O ₂ SH ⁺	Sulphur compounds	Dimethyl sulfone (methylsulfonylmethane)	2.4±0.9	1.3±0.7	<0.001	0.2±0.3	n.d.	<0.001	n.a.
95.0478	95.049	C ₆ H ₆ OH ⁺	Phenols	Phenol	6.5±4.8	3.8±2.1	<0.001	0.3±0.4	0.4±0.4	0.188	[34]
95.0854	95.086	C ₇ H ₁₀ H ⁺	Terpenes	Methylcyclohexadiene (α-terpinene fragment)	87±121	64±61	0.051	4.2±3.2	2.6±2.2	<0.001	[33]
96.0814	96.081	C ₆ H ₉ NH ⁺	Heterocyclic compounds	Ethylpyrrole	7.1±8.2	5.9±4.4	0.144	2.1±1.8	2.5±1.6	0.002	[32]
97.0282	97.028	C ₅ H ₄ O ₂ H ⁺	Aldehydes	Furfural	13±12	2.9±1.9	<0.001	1.8±2.7	0.4±0.7	<0.001	[32, 33]
97.0647	97.065	C ₆ H ₈ OH ⁺	Aldehydes/Furans	Hexadienal/ethylfuran	19±23	19±24	0.797	25±26	8.3±8.6	<0.001	[24, 32]
99.0803	99.080	C ₆ H ₁₀ OH ⁺	Aldehydes	Hexenal/methylpentenone	19±16	15±9	0.009	21±22	4.6±2.8	<0.001	[24, 29]
101.0960	101.096	C ₆ H ₁₂ OH ⁺	Alcohols	Hexenol	24±28	10±13	<0.001	13±11	3.7±4.5	<0.001	[32, 33]
103.0755	103.075	C ₅ H ₁₀ O ₂ H ⁺	Esters and acids	Methylbutanoic acid	12±9	11±11	0.414	1.1±1.4	0.7±1.1	0.002	[23]
105.0343	105.037	C ₄ H ₈ OSH ⁺	Sulphur compounds	Methional	2.0±0.8	0.4±0.4	<0.001	0.2±0.2	n.d.	<0.001	[30]
105.0689	105.070	C ₈ H ₈ H ⁺	Aromatic hydrocarbons	Styrene/ethylbenzene/vinylbenzene	13±5	1.4±0.8	<0.001	1.0±0.7	0.1±0.3	<0.001	[9]
107.0488	107.049	C ₇ H ₆ OH ⁺	Aldehydes	Benzaldehyde	33±15	13±9	<0.001	9.4±5.9	1.8±1.8	<0.001	[24, 29, 33, 35]
107.0855	107.086	C ₈ H ₁₀ H ⁺	Aromatic hydrocarbons	Xylene/ethylbenzene	58±43	43±37	0.002	9 ±16	8±15	0.398	[24, 35]
109.0658	109.065	C ₇ H ₈ OH ⁺	Phenols	Benzyl alcohol (cresol)	3.2±1.9	1.8±1.3	<0.001	0.5±0.3	0.2±0.2	<0.001	[24, 29, 33]
109.1013	109.101	C ₈ H ₁₂ H ⁺	Hydrocarbons	Cyclooctadiene	25±29	21±17	0.154	4.0±3.5	2.3±2.2	<0.001	[29]
111.0466	111.044	C ₆ H ₆ O ₂ H ⁺	Furans	Acetyl furan	3.8±3.5	2.2±1.6	<0.001	0.3±0.7	0.2±0.4	0.028	[24, 33]
111.0805	111.080	C ₇ H ₁₀ OH ⁺	Aldehydes	Heptadienal	17±20	20±24	0.213	7.3±7.2	6.4±7.4	0.120	[24, 36]
113.0960	113.096	C ₇ H ₁₂ OH ⁺	Aldehydes	Heptenal	3.3±2.5	2.6±2.2	0.010	1.6±1.3	0.7±0.7	<0.001	[24, 29, 32]
115.0738	115.074	C ₆ H ₁₀ O ₂ H ⁺	Ketones	Caprolactone	2.5±1.7	1.2±1.0	<0.001	0.2±0.1	0.1±0.1	<0.001	[29]
115.1119	115.112	C ₇ H ₁₄ OH ⁺	Ketones	Heptanone	10±10	5.5±5.9	<0.001	3.6±2.9	1.2±1.1	<0.001	[28]
121.0291	121.028	C ₇ H ₄ O ₂ H ⁺	Terpenes	cyclohexadienone (fragment)	0.8±0.4	n.d.	<0.001	0.2±0.3	n.d.	<0.001	[24, 29, 33]
121.0648	121.065	C ₈ H ₈ OH ⁺	Aldehydes	Methylbenzaldehyde - coumaran	2.5±1.2	1.1±0.6	<0.001	2.3±1.3	0.5±0.6	<0.001	[24, 32, 35]
121.1004	121.101	C ₉ H ₁₂ H ⁺	Aromatic hydrocarbons	Methylethylbenzene	8.7±6.9	7.6±6.1	0.134	0.8±0.4	0.5±0.5	<0.001	[29]
123.1170	123.117	C ₉ H ₁₄ H ⁺	Terpenes	Santene	5.6±2.3	5.9±2.3	0.242	1.2±0.5	0.7±0.4	<0.001	n.a.
127.1116	127.112	C ₈ H ₁₄ OH ⁺	Ketones	Octenone/methylheptenone	9±11	7.3±6.2	0.030	2.6±2.3	1.2±1.1	<0.001	[24, 29]
129.0901	129.091	C ₇ H ₁₂ O ₂ H ⁺	Esters and acids	Hexenyl formate	1.7±1.1	0.8±1.1	<0.001	0.4±0.3	n.d.	<0.001	[32]

129.1276	129.127	C ₈ H ₁₆ OH ⁺	Ketones	Octanone/Dimethylcyclohexanol	2.9±2.7	2.1±1.8	0.002	1.2±1.1	0.4±0.3	<0.001	[32, 33]
131.1069	131.107	C ₇ H ₁₄ O ₂ H ⁺	Esters and acids	Heptanoic acid/hexyl formate	9.9±9.4	8.2±11.0	0.138	2.3±2.1	1.1±1.7	<0.001	[24, 32, 33]
135.1170	135.117	C ₁₀ H ₁₄ H ⁺	Aromatic hydrocarbons	Methylpropylbenzene	12±14	6.2±4.9	<0.001	1.0±1.0	0.4±0.4	<0.001	[29]
136.1212	136.112	C ₉ H ₁₃ NH ⁺	Heterocyclic compounds	Butyl-pyridine/ethyl-propylpyridine	3.7±4.9	2.3±2.3	0.005	0.2±0.2	n.d.	<0.001	n.a.
137.1321	137.133	C ₁₀ H ₁₆ H ⁺	Terpenes	Various monoterpenes	368±548	252±277	0.033	13±14	5.6±7.5	<0.001	[29, 32, 33, 36]
139.1124	139.112	C ₉ H ₁₄ OH ⁺	Aldehydes	Nonadienal	27±16	19±12	<0.001	2.8±1.5	1.3±0.9	<0.001	[29, 33, 36]
141.1271	141.127	C ₉ H ₁₆ OH ⁺	Aldehydes	Nonenal	3.2±1.5	2.5±1.2	<0.001	0.9±0.6	0.5±0.3	<0.001	[29, 36]
143.1435	143.143	C ₉ H ₁₈ OH ⁺	Ketones/Aldehydes	Nonanone/nonanal	0.9±0.9	0.6±0.4	0.002	0.8±0.8	0.3±0.3	<0.001	[24, 29]
151.1114	151.112	C ₁₀ H ₁₄ OH ⁺	Terpenes	Carvacrol/ safranal	1.0±0.4	0.9±0.5	0.365	0.3±0.13	0.2±0.1	<0.001	[24, 33, 36]
153.0550	153.055	C ₈ H ₈ O ₃ H ⁺	Aldehydes	Vanillin, methyl salicylate	6.1±4.0	n.d.	<0.001	4.2±3.4	n.d.	<0.001	[33, 36, 37]
153.1272	153.127	C ₁₀ H ₁₆ OH ⁺	Aldehydes	Decadienal	29±26	4.4±2.9	<0.001	2.9±1.9	0.6±0.5	<0.001	[36]
155.1430	155.143	C ₁₀ H ₁₈ OH ⁺	Alcohols	Linalool/ geraniol	2.1±1.5	1.1±1.5	<0.001	0.6±0.4	0.2±0.3	<0.001	[29, 33, 35, 36]
171.1332	171.138	C ₁₀ H ₁₈ O ₂ H ⁺	Terpenes	Linalool oxide	2.9±2.6	0.2±0.1	<0.001	0.2±0.2	n.d.	<0.001	[24, 29, 36]
193.1587	193.159	C ₁₃ H ₂₀ OH ⁺	Terpenes	B-ionone	0.4±0.2	0.3±0.2	0.141	0.2±0.2	0.2±0.2	0.626	[29, 33, 35, 36]
195.0879	195.088	C ₈ H ₁₀ N ₄ O ₂ H ⁺	Ketones	Caffeine	n.d.	0.2±0.1	<0.001	0.2±0.1	n.d.	<0.001	[29]

468

469n.a.: Not available, n.d. : Not detected

470

471 **Table 3.** Confusion matrices showing the origin separation of black and green tea infusions for leave-group-out cross validation obtained by random
472 forests (RF), penalized discriminant analysis (PDA), support vector machine (SVM) and discriminant partial least squares (dPLS) classification models

Classification method	Black tea					Average Error rate (%)	Green tea					Average Error rate (%)
RF		Chi	Ind	Sri	Nep	32.4		Chi	Ind	Jap	Kor	39.6
	Chi	58	19	0	1		Chi	74	6	4	0	
	Ind	9	146	0	1		Ind	15	0	9	0	
	Sri	0	24	0	0		Jap	10	3	42	5	
	Nep	12	35	0	1		Kor	16	0	8	0	
PDA		Chi	Ind	Sri	Nep	36.5		Chi	Ind	Jap	Kor	45.3
	Chi	55	11	4	8		Chi	69	2	10	3	
	Ind	2	108	30	16		Ind	5	5	14	0	
	Sri	0	19	5	0		Jap	17	0	28	15	
	Nep	0	24	0	24		Kor	5	0	16	3	
SVM		Chi	Ind	Sri	Nep	37.5		Chi	Ind	Jap	Kor	42.2
	Chi	57	16	1	4		Chi	62	5	14	3	
	Ind	4	105	22	35		Ind	7	14	3	0	
	Sri	0	18	6	0		Jap	17	0	31	12	
	Nep	1	26	0	21		Kor	8	0	12	4	
dPLS		Chi	Ind	Sri	Nep	43.3		Chi	Ind	Jap	Kor	48.4
	Chi	44	12	7	15		Chi	67	0	12	5	
	Ind	2	101	29	24		Ind	7	4	13	0	
	Sri	0	19	5	0		Jap	17	0	24	19	
	Nep	0	27	0	21		Kor	6	0	14	4	

473

474

475

476

477