Untargeted metabolomics reveals differences between commercial and non-commercial *Camellia sinensis* cultivars used in black tea production

Christopher Nyarukowa[a], Mari van Reenen[b], Robert Koech[c], Shayne Mason[b], Zander Lindeque[b], Samson Kamunya[c], Richard Mose[d], Zeno Apostolides[a*]

[a]Department of Biochemistry, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

[b]Human Metabolomics, North-West University (Potchefstroom Campus), South Africa

[c]Kenya Agriculture and Livestock Research Organization, Tea Research Institute, 820-20200 Kericho, Kenya

[d]James Finlay (Kenya) Limited, P.O. Box 223, Kericho 20200, Kenya

*Address correspondence to:
Prof. Zeno Apostolides, Department of Biochemistry, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa. Email: zeno.apostolides@up.ac.za. Phone: +27(0)12 420-2486.

Abstract

Tea (*Camellia sinensis*) has enthralled both consumers and researchers, and its popularity has increased, due to its taste, aroma and its medicinal attributes, owed largely to its metabolites. The catechins and theaflavins in green and black tea, respectively, have been documented to possess anti-oxidant, anti-inflammatory, anti-cancer, as well as cardiovascular disease-preventing properties. Tea consumers concern themselves with the quality of tea, in particular, its taste and aroma based on which consumers are willing to pay premium prices for the best quality teas. In turn, the quality of tea is undeniably affected by variations in its metabolite composition. In this study, two groups of black tea cultivars, one commercial and the second non-commercial, were compared using a metabolomics approach. Data were generated via untargeted GC-MS and semi-targeted ^{1}H-NMR. The GC-MS results differentiated between the two groups, based on arabinose, sucrose, phloroglucinol and xylonic acid. The ^{1}H-NMR results differentiated between the two groups, based on caffeine, catechin, EC, EGC, and the amino acids alanine, isoleucine, leucine, theanine, and valine. These metabolites applicability in the discrimination of newly developed cultivars into potentially commercialisable and non-commercialisable groups at an early stage in the tea improvement programme is demonstrated. This information may also help tea breeders to select promising high quality black and drought tolerant improved tea cultivars either for release or further field evaluations.

Keywords

Camellia sinensis; catechin; metabolomics; tea quality

1. Introduction

Tea has been documented to be the most consumed non-alcoholic beverage worldwide, second only to water (Cabrera et al., 2006). According to Dutta et al. (2011), tea production has increased from 850 000 tonnes in 2003 to 980 000 tonnes in 2007 (Dutta et al., 2011) and to 2 414 802 tonnes in 2018 (www.reportlinker.com/tea/reports). Tea quality is undeniably affected by variations in its metabolite composition, which determine its commercial market value (Qin et al., 2013). According to Le Gall et al., (2004), the taste of green tea is determined by the cultivar of tea tree, the season of plucking, as well as the method of cultivation employed. Amino acids such as theanine, which makes up between 60-70% of the free amino acids found in tea leaves, are responsible for tea's brothy taste. Theanine also has anxiety and stress relieving properties (Unno et al., 2013). While its astringent taste can be attributed to catechin levels and lastly its bitter taste attributed to caffeine (Pongsuwan et al., 2007). Caffeine also has fatigue relieving effects (Lorist and Tops, 2003). This fatigue relieving effect of caffeine is probably the most important attribute of tea since antiquity. Tea leaf comprises about 10% of total catechins and 2 - 3% (dry weight) caffeine. The predominant catechins are catechin (CAT), epicatechin (EC), epicatechin gallate (ECg), epigallocatechin (EGC), and epigallocatechin gallate (EGCg) (Koech et al., 2018), with EGCg being the principal catechin accounting for approximately 50-75% (Zhou et al., 2019) of all the total catechins. Black tea, on the other hand, consists of theaflavins and thearubigins as its major phenolics, which result from the dimerisation and polymerisation of the catechins. Black tea consists of theaflavin (TF1), theaflavin-3-monogallate (TF2), theaflavin-3′-monogallate (TF3), and theaflavin-3,3′-digallate (TF4). Tea's popularity as a beverage is dependent on its flavour, comprising of taste and aroma. Non-volatile organic compounds are responsible for its taste, while volatile organic compounds are responsible for its aroma (Chaturvedula and Prakash, 2011). Volatile organic compounds in tea fall into one of two groups, with Group I comprising of non-terpenoids, such as hexanol which confers the fresh green aroma; and Group II comprising of terpenoids, responsible for its sweet flowery aroma (Rawat et al., 2007). High-quality black teas are rich in Group II compounds and due to their flowery nature, achieve significantly higher prices; e.g. high-quality varieties of Longjing tea (pan-roasted green tea) sell for USD 16.5, while the low-quality varieties sell for USD 1 per Kg (Yu et al., 2008).

Sensory evaluation by expert tea tasters has traditionally been employed to establish its specific aroma profile, which is important in determining the tea quality grade (Schuh and

Schieberle, 2006). These trained tasters have developed a language of their own which they use to describe numerous characteristics of a tea brew. This sensory evaluation has its own deficiencies, such as being time-consuming, prone to human subjectivity and inherent variability between tasters (Group et al., 2011). In this paper, we consider a more objective method for determining the quality of tea using a broader metabolomics approach to explore the metabolite composition of two groups of tea cultivars. Metabolomics has been defined as "*the study of the quantitative measurement of the dynamic multi-parametric metabolic response of biological system and changes in metabolite concentrations or fluxes related to genetic or environmental perturbations*" (Dunn et al., 2011). It is a discipline which assesses, classifies and quantifies endogenous and exogenous metabolites in a variety of biological samples. The information obtained from the study of metabolites is crucial in that it informs scientists about a biological system's functional state, explaining the organism's phenotypic traits. The analysis of metabolite profiles in biological systems allows for the unbiased analysis of alterations in their metabolic status as a result of e.g. genetic composition and environmental stress (Ludwig and Viant, 2010).

It has been reported that areas with higher rainfall tend to produce teas of inferior quality and that black teas from South India have elevated concentrations of aroma causing metabolites during the dry season as compared to the rainy season. In these studies, 40 or fewer compounds were used to classify tea quality (Kowalsick et al., 2014). The profiling of plant metabolites has developed into a major metabolomics field of study, the reason being that plants manufacture a wide array of metabolites. The genetic improvement of crops with metabolomics is fast becoming a popular method; this has resulted in an increased demand for plant breeders skilled in the field of metabolomics (Chugh, 2013). When developing new cultivars, crop breeders encounter a common challenge of identifying important selection criteria. Tea breeders criteria of selection include but not limited to yield, quality, and drought tolerance. *C. sinensis* is an important cash crop for many countries with China, India, Kenya and Sri Lanka being leading world producers and exporters of black tea (ITC, 2019). According to the Kenya National Bureau of Statistics (2019), tea is the largest agribusiness in Kenya, with the total export volume of January 2019 being significantly higher, at 47.92 Million Kg compared to the January 2018 total export volume of 31.94 Million Kg. It is for this reason that tea quality is an important selection criterion from an economic perspective as it is the major determinant of market price (Yan, 2007). Tea quality, whether for black, white, purple or green tea, is governed by the metabolic profile/composition of the tea leaves,

influencing its aroma, briskness, brightness and taste (Dutta et al., 2011); the agronomic traits i.e. yield and quality are dependent on leaf physiognomies. As previously noted, due to the effects of global warming, specifically altered precipitation patterns, elevated temperatures and protracted drought spells in the tea growing regions, the Kenyan tea industry has been facing challenges (FAO, 2015). It is for this reason that rigorous breeding programmes need to be developed to produce new cultivars with better metabolic profiles and improved drought tolerance. A previous study on drought tolerance in *C. sinensis* saw the development of the Short-time Withering Assessment of Probability for Drought Tolerance (SWAPDT) method. The SWAPDT method was validated by targeted metabolomics, to predict tolerance in tea cultivars by generating metabolic profiles, which showed the differences between the drought-tolerant, and drought-susceptible cultivars under wet conditions. This method employs the % relative water content of tea leaves after a five-hour withering period (Nyarukowa et al., 2016).

Nuclear magnetic resonance (^{1}H-NMR) spectroscopy and gas chromatography mass spectrometry (GC-MS) were the analytical metabolomics platforms used in this study. GC-MS, though tedious in the sample preparation stage, has a higher sensitivity as compared to ^{1}H-NMR, capable of detecting metabolites with concentrations lower than the limit of detection of ^{1}H-NMR. ^{1}H-NMR has the advantage of having simple sample preparation, as well as being semi-quantitative and non-destructive (Dumas et al., 2006), but limited by resolution and the availability of plant metabolites in compound databases. The objective of this study was to use data generated through an untargeted metabolomics approach to identify metabolites, which were expressed differently in the two groups, namely the commercial (Comm) and non-commercial (NComm) cultivars.

2. Materials and Methods

2.1. Plant material

All the cultivars used in this study were maintained using uniform agronomic practices in an experimental field site in Kericho (0° 22′ S, 35° 17′ E), Kenya as described by Koech et al., (2018). Sixty open-pollinated cultivars, pre-selected for their high yield, and good tea liquor since the 1950s formed the Comm group. These cultivars were vegetatively propagated by stem cuttings from elite mother bushes. Each Comm cultivar is cultivated in over 10 Hectares with 13,448 bushes per Ha. The NComm group of 250 cultivars were the F_1 progeny of a reciprocal cross between two heterozygous parental clones TRFK 303/577 and GW Ejulu.

The NComm cultivars comprise two populations of TRFK St. 504 (TRFK 303/577 x GW Ejulu) with 106 progeny, and the TRFK St. 524 (GW Ejulu x TRFK 303/577) with 144 progeny, which were bred at the Tea Research Institute (TRI) of Kenya (Koech et al., 2018). The GW Ejulu is a commercial cultivar that produces high-quality black tea, with high total catechins and moderate caffeine content; it is, however, a low-yielding and drought-susceptible clone. Cultivar TRFK 303/577, on the other hand, is a high yielding, drought tolerant (DT) commercial cultivar, which produces medium-quality black tea, with moderate levels of caffeine and total catechins. All the plants were vegetatively propagated and planted in 15-bush observation plots comprising 3 rows and 5 plants per row spaced at 1.22 m between rows and 0.61 m within rows (i.e. 13,448 plants/ha) in a randomised complete block design with three replicates.

2.2. Sample collection and processing

About 500 grams fresh shoots comprising two leaves and a bud were harvested from the respective tea bushes, between September 2013 and February 2014. The fresh shoots were placed in appropriately labelled zip-lock plastic bags (Nyarukowa et al., 2018). The plastic bags were placed on ice blocks to keep cool until processing at the TRI mini-factory within 24 hours. Half the shots of each sample were freeze-dried and ground using a coffee grinder, sieved using a 355 µm sieve, sealed in zip-lock plastic bags and stored at 4°C in a fridge until analysis.

2.3. GC-MS sample preparation and analysis

2.3.1. Sample preparation

A 70% MeOH solution was used for extraction. For all samples of approximately 150 mg, 1.5 mL extraction solution was added. The samples were vortex mixed and incubated for 10 minutes at 70°C. The samples were vortex mixed halfway through the incubation period as well as at the end. After cooling, the samples were centrifuged for 5 minutes at 6000 g and the 1 mL supernatant transferred to GC vials before drying under nitrogen. The dried samples were derivatised by adding 120 µl methoxyamine (10 mg/mL in pyridine) and incubated for 1 hour at 60°C; followed by the addition of 80 µl BSTFA (containing 1% trimethylchlorosilane) and incubated for another hour at 60°C. Samples were transferred to inserts before GC-MS analysis. Pooled quality control (QC) samples were prepared, and these underwent the same extraction and derivatisation procedures as the samples.

2.3.2. GC-MS analyses

Analyses were performed on a GC-TOF-MS system, comprising of an Agilent 7890A GC front-end system with an Agilent 7693 autosampler and a Leco Pegasus HT TOFMS. Hydrogen was used as carrier gas at a flow-rate of 1.8 mL/min; 0.2 µl sample was injected in splitless mode (allowing 30s purge delay). The inlet temperature was kept at 250°C. Compounds were separated on a Restek RX-1MS column (20 m x 180 µm x 0.18 µm). The transfer line and source temperatures were 250 and 200°C, respectively. Solvent delays of 200 s were allowed where after masses (50 – 800 m/z) were recorded at 20 spectra/sec. Universal EI settings were used for ionisation while the detector was operated at 50 V above tune voltage.

2.4. ^{1}H-NMR sample preparation and analysis

2.4.1. ^{1}H-NMR buffer solution

A 1.5 M KH_2PO_4 buffer solution was prepared by dissolving 20.4 g of KH_2PO_4 in 80 mL of deuterium oxide (D_2O). Next, 13 mg of sodium azide and 100 mg of trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) were dissolved in 10 mL of D_2O and added to KH_2PO_4 solution. The combined solution was mixed well under sonication before adjusting the pH to 7.4 using potassium hydroxide in H_2O. The final solution was then transferred to a 100 mL volumetric flask and the volume topped up to the mark using D_2O.

2.4.2. ^{1}H-NMR sample preparation

Freeze-dried samples were sent in individual plastic bags of 50 mg weight to the ^{1}H-NMR lab. A pooled QC sample was created by collecting 5 mg from each of n=294 samples. Samples were prepared by adding 4.5 mL ddH_20 to each 45 mg weight of the dry sample to create a 10 mg/mL concentration. Each sample was vortexed at 0, 20 and 40 minutes. At 60 minutes, a volume of 540 µL of the sample was collected in a microcentrifuge tube, with 60 µL ^{1}H-NMR buffer solution. The sample was mixed under vortex and centrifuged at 12 000 *g* for 5 minutes to sediment any particulates. A final volume of 540 µL of supernatant was carefully transferred to a 5 mm ^{1}H-NMR glass tube and loaded onto an autosampler for ^{1}H-NMR analysis.

2.4.3. ^{1}H-NMR analyses

The samples were measured at 500 MHz on a Bruker Avance III HD NMR spectrometer equipped with a triple-resonance inverse (TXI) ^{1}H{^{15}N, ^{13}C} probe head and x, y, z gradient coils. ^{1}H spectra were acquired as 128 transients in 64 K data points with a receiver gain of 64 and a spectral width of 10 000 Hz. The sample temperature was maintained at 300K and

the H_2O resonance was presaturated by single-frequency irradiation during a relaxation delay of 4 s, with a 90° excitation pulse of 8 μs. Shimming of the sample was performed automatically on the deuterium signal. The resonance line widths for TSP and metabolites were <1 Hz. Fourier transformation and phase and baseline correction were done automatically. Software used for ^{1}H-NMR processing was Bruker Topspin (V3.5). Bruker AMIX (V3.9.14) was used for metabolite identification and quantification. (Ellinger et al., 2013).

2.5. Metabolite identification

Spectral matching to the NIST11 commercial library (for GC-MS metabolites) and Bruker BBIOREFCODE (pH 7.0) and in-house pure compound spectral libraries (pH 7.4) (for ^{1}H-NMR metabolites) were used to identify the compounds. A level 2 identity was awarded when a spectral match of 80% similarity was achieved. A level 1 identity was awarded when the retention time or retention index of the GC-MS information matched that of standards (Schymanski et al., 2014) or 2D ^{1}H-NMR information confirmed 1D ^{1}H-NMR spectral identifications.

2.6. Data pre-processing

ChromaTOF software (Leco) was used to perform data extraction for the GC-MS data, which included baseline removal using the "spanning" tracking method, with an offset of 1. The software performed automatic smoothing. An expected peak width of 3 s and signal-to-noise ratio of 20 was used to detect the peaks with five apexing masses. GC-MS data was normalised using the "total useful signal" correction method, as described by Warrack et al., (2009). A subset of the data was aligned for exploratory statistical analysis since the add-on function of ChromaTOF (Statistical Compare) is unable to align > 250 samples. The subset consisted of approximately 140 randomly selected Comm and NComm cultivar samples from every batch (including QCs). With the exploratory statistics, a list of compounds that differed between the groups was generated, which was used to create a reference chromatogram within ChromaTOF. The reference was used to extract the peaks of interest from all the samples in a "targeted" manner. The target peaks lists were aligned into a data matrix with MS Excel using the "consolidate" function. The reason for pre-processing was to transform the data to enhance ease and improve the data analysis. ^{1}H-NMR spectra pre-processing involved binning and scaling (Ebbels et al., 2011).

^{1}H-NMR variables were scaled relative to the internal standard (TSP) by dividing each bin by the corresponding TSP value for the same sample. Next, the combined GC-MS and ^{1}H-

NMR variables with more than 10% missing values in both groups were eliminated; if two variables had a high correlation, one was removed; outliers were removed. The remaining missing values for each group, deemed to be below the quantification threshold of the instrument, were imputed with random numbers drawn from a uniform distribution between one and two-thirds of the lowest non-zero observations. Imputations were performed for each variable independently.

2.7. Statistical analysis

In metabolomics, uni- and multivariate statistical techniques are used in combination to help pinpoint relevant variation (e.g. between groups of interest) in datasets that are often large and high-dimensional (Saccenti et al., 2014). The univariate statistical tool used here was the t-test with resulting p-value and associated effect size. Independent samples t-tests were performed assuming unequal group variance and a 5% significance using MATLAB with Statistics Toolbox (2019), version 9.5.0 (R2018b) software (Natick, Massachusetts: The MathWorks Inc). Effect sizes were incorporated as an indication of the practical relevance of significant differences ($P \leq 0.05$) based on Cohen's d-value (Ellis and Steyn, 2003) and calculated manually as the absolute difference between group means divided by the larger of the two group standard deviations (SD). To control the family wise error rate, p-values were adjusted using the Bonferonni-Holm correction for multiple testing (Holm, 1979)).

Two multivariate approaches were included, principal component analysis (PCA), and partial least squares discriminant analysis (PLS-DA). PCA was used as the first step in multivariate statistical analysis, as it is highly expedient when it comes to e.g. outlier identification, pattern and trend detection. PCA scores plots were generated to provide a visual summary of the predominant variation in each dataset and the association with the two experimental groups. This could be achieved as the PCA models constructed here were unsupervised and so received no group information. PLS-DA is another multivariate statistical approach employed in metabolomics data analysis. PLS-DA has been described as a versatile algorithm capable of being used for discriminative variable selection, as well as descriptive and predictive high-dimensional dataset modelling (Worley et al., 2013). PLS-DA is a better suited statistical approach, compared to the PCA, when it comes to distinguishing between the two groups of samples as it is a supervised method, especially in instances where the metabolite profiles are influenced/affected by several factors (Kang et al., 2008). That said, PLS-DA models are prone to overfit and must be validated. A leave-one-out cross-validation (LOO-CV) procedure was followed here to validate the variance explained in the grouping

variables. PLS-DA scores plots were generated to assess the predictive ability of the model. Prior to identifying compounds that are largely responsible for any visible separation, the goodness-of-fit statistics (R-squared), as well as the LOO-CV (Q-squared) statistics, were compared to assess model validity. R-squared values above 80% were considered sufficient but conditioned to no dramatic deterioration during LOO-CV, that is Q-squared values above 60% were considered acceptable. PCA and PLS-DA analysis were performed using MATLAB with Statistics Toolbox (2019), version 9.5.0 (R2018b) software (Natick, Massachusetts: The MathWorks Inc) in conjunction with the PLS_Toolbox (2019), version 8.7 software (Wenatchee, WA: Eigenvector Research Inc. Software available at http://www.eigenvector.com). Prior to statistical analysis, the data were pre-processed to help ensure the accuracy of results. The GC-MS analysis was untargeted and as expected the resulting dataset contained many zero values. Both the ^{1}H-NMR as well as GC-MS datasets were log transformed (shifted natural log transformation with shift parameter set to 1) to correct for the skewness in distribution known to plaque metabolomics data, and auto-scaled (subtracting the mean and dividing by the SD) so compounds in different abundances receive equal attention during multivariate analysis (van den Berg et al., 2006). Finally, the PCA model based scores plots and Hotelling's T-squared distances were employed to detect outliers within each group given a 95% confidence interval (CI).

3. Results

3.1. GC-MS and ^{1}H-NMR PCA

Results presented in Figure 1 show that the ellipsoids representing 95% CI of score centroids of the Comm and NComm groups separate better by ^{1}H-NMR (Figure 1B) than GC-MS (Figure 1A). The percentage of the overall variation in the measured compounds explained by each principal component (PC) is indicated along the three axes.

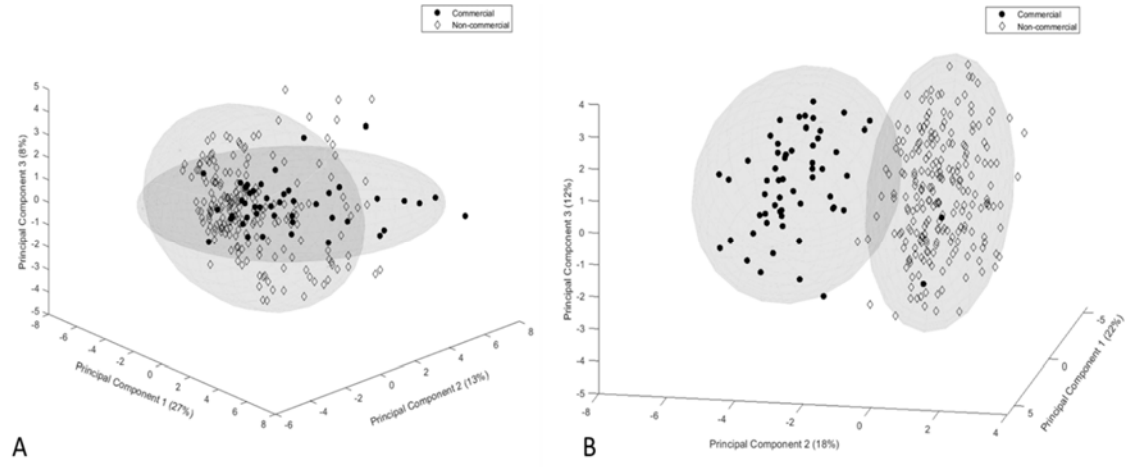

Figure 1: The 3D PCA scores plots for PCs one, two and three, for GC-MS (A) and ^{1}H-NMR (B), the two parents of the NComm cultivars are represented by the two black circles in the ellipsoid of the NComm cultivars. The sum of the three PCs explain 48% and 52% of the variation found in the GC-MS and the ^{1}H-NMR respectively.

3.2. GC-MS and ^{1}H-NMR PLS-DA

Similar to PCA plots, ellipsoids represent 95% CI of score centroids of each group. The percentage of the overall variation in the measured compounds (X) and group membership (Y), as explained by each latent variable (LV), which is indicated along each axis, as shown in Figure 2.

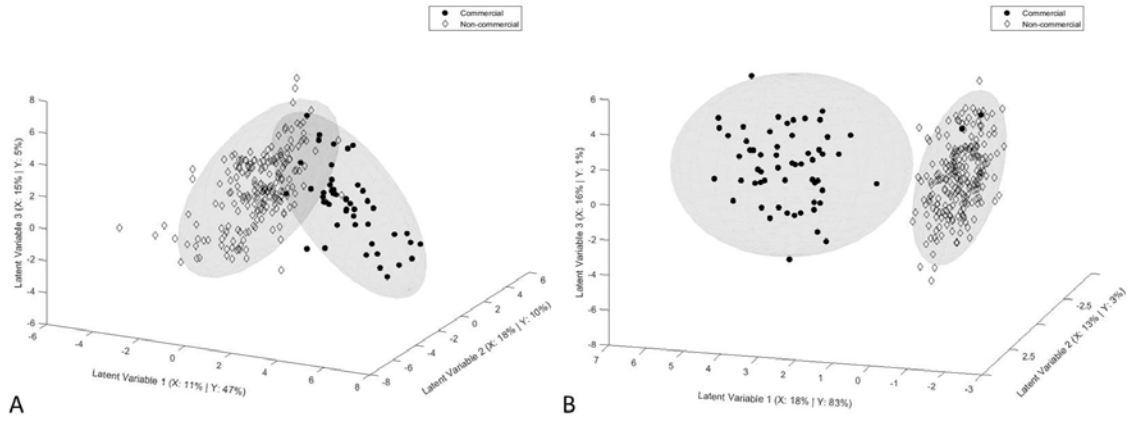

Figure 2: The 3D PLS-DA scores plots for LVs one, two and three, for GC-MS (A) and ^{1}H-NMR (B), the two parents of the NComm cultivars are represented by the two black circles in the ellipsoid of the NComm cultivars. The goodness-of-fit values achieved for the GC-MS model were R^2=62% and Q^2=55% making it unreliable for discriminant identification, while the ^{1}H-NMR model was deemed reliable with R^2=87% and Q^2=85%.

Table 1: The list of tentatively identified metabolites detected by GC-MS, expressed in arbitrary units.

Comm vs NComm variables	Analytical platform	Relative normalised intensity		Fold change	Reported literature concentration (mg/g)	References
		Comm	NComm			
Acetoacetic acid	GC-MS	0.050	0.040	1.2	20.0	Naveed et al., 2017
Arabinose	GC-MS	0.011	0.005	2.2*	37.3	Nara et al., 2001
Arabitol	GC-MS	0.010	0.008	1.3	20.0	Roser et al., 1992
Catechin	GC-MS	0.070	0.110	0.6*	30.0	Gramza et al., 2006
1-Cyclohexenecarboxylic Acid	GC-MS	0.034	0.027	1.3*	4.7	Baeza et al., 2016
Gallic acid	GC-MS	0.056	0.040	1.4	5.4	Kaneko et al., 2006
Glycerol	GC-MS	0.006	0.006	1.0	10.0	Jones et al., 2008
Linolenic acid	GC-MS	0.002	0.001	2.0	0.1	Stephan and Steinhart, 2000
Malic acid	GC-MS	0.011	0.008	1.4*	10.0	Ackerman et al., 1992
Phloroglucinol	GC-MS	0.003	0.003	1.0	45.1	Matanjun et al., 2008
Psicose	GC-MS	0.0005	0.001	0.5*	1.3	Mu et al., 2012
Ribitol	GC-MS	0.007	0.003	2.3*	20.0	Roser et al., 1992
Sucrose	GC-MS	0.040	0.012	3.6*	30.9	Kumar et al., 2011
Threonic acid	GC-MS	0.006	0.018	0.3*	0.2	Horemans et al., 2000
Xylonic acid	GC-MS	0.001	0.001	1.0	4.5	Habibi et al., 2004
Total sweeteners	GC-MS	0.095	0.047	2.0*	N.A	N.A

*indicate a statistically significant difference in the mean concentration of the metabolite between the Comm and NComm cultivars at the 95% level of significance after correcting for multiple testing. N.A = not available.

Table 2: The list of metabolites detected by ^{1}H-NMR, expressed in mg/g.

Comm vs NComm variables	Analytical platform	Comm concentration (mg/g)			NComm concentration (mg/g)			Fold change	Reported literature concentration (mg/g)	References
		Min	Max	Mean	Min	Max	Mean			
Acetic acid	^{1}H-NMR	25.01	60.97	34.8	28.80	70.71	42.2	0.8	40.0	Bandurski and Schulze, 1977
Alanine	^{1}H-NMR	0.11	0.32	0.2	0.16	0.40	0.2	1.0	4.2	Min et al., 2017
Caffeine	^{1}H-NMR	6.24	20.02	12.6	6.14	19.12	11.6	1.1*	30.0	Chin et al., 2008
Catechin	^{1}H-NMR	6.17	30.32	15.4	2.66	28.89	14.2	1.08	30.0	Gramza et al., 2006
Chlorogenic acid	^{1}H-NMR	2.65	6.44	4.2	2.51	6.07	4.0	1.02	6.9	Marks et al., 2007
Epicatechin	^{1}H-NMR	7.92	28.16	14.6	8.84	26.26	14.2	1.02	70.0	Gramza et al., 2006
Epicatechin gallate	^{1}H-NMR	4.01	23.03	13.5	4.32	22.13	13.2	1.02	169.0	Gramza et al., 2006
Epigallocatechin	^{1}H-NMR	20.55	117.41	51.7	17.12	113.49	49.6	1.04*	150.0	Gramza et al., 2006
Epigallocatechin gallate	^{1}H-NMR	19.02	66.6	39.0	17.36	57.58	35.6	1.1*	173.0	Gramza et al., 2006
Formic acid	^{1}H-NMR	19.07	46.40	27.2	16.01	31.08	21.0	1.3*	21.0	Sanhueza, E., and Andreae, 1991
Gallic acid	^{1}H-NMR	0.53	2.04	1.1	0.60	2.98	1.0	1.1	5.4	Kaneko et al., 2006
Glucose	^{1}H-NMR	5.38	11.72	7.5	4.97	10.36	7.2	1.04	6.9	Melgarejo et al., 2000
Isoleucine	^{1}H-NMR	0.13	0.41	0.2	0.01	0.27	0.2	1.0	2.6	Min et al., 2017
Leucine	^{1}H-NMR	0.06	0.23	0.2	0.08	0.30	0.1	2.0*	3.9	Min et al., 2017
Methanol	^{1}H-NMR	0.07	0.29	0.2	0.07	0.25	0.1	2.0*	0.04	Fall and Benson, 1996
Sucrose	^{1}H-NMR	5.91	21.54	15.0	6.61	23.14	13.6	1.1*	30.9	Kumar et al., 2011
Theanine	^{1}H-NMR	2.82	22.22	8.6	4.19	13.79	8.0	1.1	50.0	Vuong et al., 2011
Quinic acid	^{1}H-NMR	1.01	2.70	1.9	1.03	3.04	2.0	1.0	5.0	Rodrigues et al., 2007
Valine	^{1}H-NMR	0.11	0.29	1.7	0.09	0.27	1.7	1.0	3.4	Min et al., 2017
Total amino acid	^{1}H-NMR	13.22	33.89	21.8	11.97	32.85	20.9	1.04	N.A	N.A
Total catechins	^{1}H-NMR	55.73	241.18	121.58	51.22	160.3	114.96	1.1*	N.A	N.A
Total sweeteners	^{1}H-NMR	4.37	22.91	9.3	3.51	15.81	8.7	1.07*	N.A	N.A

*indicate a statistically significant difference in the mean concentration of the metabolite between the Comm and NComm cultivars at the 95% level of significance after correcting for multiple testing. N.A = not available.

4. Discussion

Metabolomics statistical data analysis can be employed as a supportive tool to aid breeders in the selection and improvement process for new tea cultivars. As mentioned, the objective of this study was to identify potential classifiers for the 310 genotypes investigated into either of the two groups, Comm or NComm cultivars, with GC-MS and ^{1}H-NMR as the metabolomics platforms. Figures 1 and 2 show the PCA and PLS-DA plots, respectively, considering the clustering of the Comm vs the NComm cultivars. From these plots, it can be seen that there is some separation between the two groups in each platform. Table 1 and 2 show a total of 15 and 19 metabolites, which were identified using GC-MS and ^{1}H-NMR respectively. The GC-MS results showed that arabinose, catechin, gallic acid, glycerol, phloroglucinol, sucrose and xylonic acid, a sugar acid generated through the complete oxidation of xylose, were detectable metabolites, which separated the Comm from the NComm cultivars in terms of arbitrary response units. Arabinose, sucrose and xylonic acid were higher in the Comm cultivars. These three metabolites have been shown to have positive correlations with metabolites such as ribose and sucrose, in other studies. They have been shown to play a role in improving the sweet taste of sugarcane (Lavarack et al., 2002) and to be up-regulated during strawberry fruit maturation (Zhang et al., 2010). It can thus be postulated that the tea cultivars with high levels of these sugars will produce sweet-tasting, higher-quality liquor. Arabinose has been reported in a study on serendipity berries from the *Dioscoreophyllum cumminsii Diels* plant, which is indigenous to tropical West Africa, and is grown in Guinea, Cameroon, and in the rain forests of central Africa. Arabinose levels were found to be higher in the fruits of some varieties of these berries and were determined to be the reason for the sweetness in these varieties (Inglett, 1971). In addition, polyols such as arabitol and ribitol, which also enhance the sweetness of fruits, have been reported in the literature at concentrations ranging between 20 and 60 mg/g dry weight (Roser et al., 1992). In a study evaluating the sucrose, and taste-related amino acids content in soybean, a sucrose concentration of 30.9 mg/g dry weight was reported (Kumar et al., 2011). The present study, however, detected average sucrose levels of 15 mg/g and 13.6 mg/g dry weight for the Comm and NComm cultivars respectively. Our results for higher concentrations of arabinose and sucrose in the Comm cultivars, agree with those in the literature where these compounds are higher in sweeter sugarcanes, ripening strawberries and sweeter cultivars of serendipity berries.

Literature has shown that abiotic stress such as drought affects the photosynthetic pathway of plants (Kerchev et al., 2012), and in so doing drastically impacts their primary metabolism, which in turn affects sugars, sugar alcohols, and amino acids. The DT plants, tend to be of a higher quality than the drought susceptible (DS), as they efficiently up-regulate their production of sugars, which they utilise as an energy source during stress (Kaplan and Guy, 2004). This could also explain why the DT cultivars produce better tasting liquor. Further, carbohydrates have also been shown to influence the biosynthesis of other energy-generating metabolites, responsible for the alteration of gene expression and signal transduction (Hoekstra et al., 2001). Phloroglucinol is a plant polyphenolic compound, which possesses antioxidant properties. It has been compared to ascorbic acid, and has been shown to be more powerful against e.g. DPPH and peroxide radicals; it is considered a natural antioxidant (Adkins et al., 2007; Archana and Vijayalakshmi, 2018). Phloroglucinol has been described in the literature as sweet (Taylor, 1928), and contributes to the sweet fruity taste of the grapes used to make Pinot noir wines (Cortell et al., 2008). Phloroglucinol was one of the metabolites identified, which was higher in the Comm cultivars as compared to the NComm cultivars. The higher concentration of phloroglucinol in the Comm cultivars agrees with the higher concentration of this compound in sweet fruity Pinot noir grapes. The higher concentration of phloroglucinol in the Comm cultivars can be explained in part by the fact that DT cultivars have been shown to have higher levels of polyphenols, which in turn results in them having a higher levels of antioxidants so they are able to scavenge free radicals better than the DS cultivars and this results in their survival under drought stress (Nyarukowa et al., 2016). Malic acid is a dicarboxylic acid, which was identified as a distinguisher between the Comm and NComm cultivars. This compound has been reported in ripening apples at concentrations of 10 mg/g (Ackermann et al., 1992); it has been reported to be responsible for the sour taste of fruits. In another study on apples by Ma et al., (2015), where the sugar and malic acid composition in cultivated vs wild apples was compared, it was found that a significant difference between the malic acid concentrations of the cultivated vs the wild apples existed. Furthermore, the study also showed that malic acid composition highly correlated with that of glucose and sucrose contents, suggesting that the selection of fruit acidity also has a significant effect on the amounts of sugars present in apple fruits. This means sugar metabolism is influenced by malic acid accumulation (Borsani et al., 2009). The wild apples were shown to be more acidic as compared to the cultivated apples. Apple breeders select for apples richer in malic acid content, due to its strong impact on sugar concentration in apples, resulting in sweeter apples (Ma et al., 2015). The Comm tea cultivar

results for malic acid agree with the malic acid results found in literature for commercial apple cultivars (Ma et al., 2015). Psicose was also detected in this study; this metabolite has been documented to confer sweetness to, for example, Worcester sauce and fruit juice. Oshima et al., (2006) reported psicose, a product of fructose breakdown, at concentrations ranging from 0.005 mg/g in coffee to 1.31 mg/g in Worcester sauce. Our study found the arbitrary psicose units to be significantly lower in the Comm cultivars compared to the NComm cultivars, indicating it is a significant distinguisher.

Some key metabolites responsible for the taste of tea are caffeine, which comprises up to 5% of the shoot dry weight; theobromine and theophylline, which are < 3% of the shoot dry weight (Khan and Mukhtar, 2007). Caffeine has, in addition to being a stimulant, been documented to contribute to tea briskness (Nitin Seetohul et al., 2006), while theophylline and theobromine have been shown to contribute to the mellowness and sweetness of oolong tea (Chaturvedula and Prakash, 2011). Furthermore, research has shown that caffeine reduces weight gain by suppressing the accumulation of body fat (Kobayashi-Hattori et al., 2005). The ^{1}H-NMR results indicate that levels of caffeine were higher in the Comm cultivars as compared to the NComm cultivars. According to Chin et al., (2008), the average caffeine content in black, and green tea is 7-30 mg/g serving of tea, with the average size of a tea bag being 2 g. The current untargeted study found average green leaf caffeine content of 12.6 and 11.6 mg/g in the Comm and NComm cultivars, respectively. These results agree with those of Mazzafera and Silvarolla, (2010), which showed that coffee beans from high-quality cultivars had a higher caffeine concentration of 25 mg/g dry weight compared to the low-quality cultivars with a lower caffeine concentration of 9.64 mg/g.

The ^{1}H-NMR results show that amino acids valine and isoleucine were detected across all samples. These two amino acids were found to be higher in the Comm cultivars. A study on Poplar trees further showed that isoleucine levels were up-regulated in the DT Poplar trees than in the DS under drought stress (Hamanishi et al., 2015). Moreover, amino acids have been documented to improve taste and aroma in tea infusions, namely alanine, leucine, phenylalanine, tryptophan, tyrosine, and valine (Tan et al., 2011). Alanine and phenylalanine have been reported as being responsible for the flowery and rose-like aromas of tea, respectively, whilst leucine has been shown to produce a spicy aroma (Sanderson and Grahamm, 1973).

Chlorogenic acid was one of the metabolites consistently detected in the Comm and NComm cultivars. Chlorogenic acid has been shown to contribute to bitterness (Iiyama et al., 1995) in

foods such as coffee. In a study by Szejtli and Szente, (2005), chlorogenic acid was complexed with the tasteless β-cyclodextrin, to eliminate the undesired bitter taste and resultant sensation in the mouth. This elimination of chlorogenic acid from coffee extracts using β-cyclodextrin was as efficient as using activated charcoal (Imamura et al., 1995). In another study on apples, it was shown that from the 20 varieties investigated, the sweeter varieties were those with lower chlorogenic acid concentrations, as low as 1.8 mg/g fresh weight, with the less sweet varieties having as high as 6.9 mg/g fresh weight (Marks et al., 2007). In the present study, chlorogenic acid concentrations were comparable in both groups with the Comm and NComm cultivars having average dry weight concentrations of 4.2 and 4 mg/g, respectively.

Quinic acid was also detected in the Comm and NComm cultivars. Literature has documented quinic acid to be responsible for the astringency taste associated with coffee (Buffo and Cardelli-Freire, 2004). However, there was no statistically significant difference between the Comm and NComm cultivars for quinic acid. The metabolite 1-cyclohexenecarboxylic acid was also detected in the present study, and was significantly higher in the Comm cultivars than the NComm cultivars. Cinnamic acids are trans-phenyl-3-propenoic acids found in plants as e.g. p-coumaric, caffeic, ferulic, dimethoxycinnamic, and trimethoxycinnamic acids, to name a few. These metabolites conjugate, through their carboxylic groups, with amino acids, polysaccharides, and glycosides. The most predominant of these reactions is the transesterification with quinic acid to form cinnamate esters, which are collectively known as chlorogenic acids (Baeza et al., 2016). Furthermore, cinnamic acids conjugated with a derivative from quinic acid, shikimic acid, results in 3,4,5-trihydroxy-1-cyclohexenecarboxylic acid, which is also a cinnamate ester (Jaiswal et al., 2010). These cinnamate esters are extensively found in fruits and vegetables, particularly green coffee beans, which are rich in this type of polyphenol, up to 14% (w/w) (Clifford, 2000). 1-cyclohexenecarboxylic acid, like other chlorogenic acids, is bitter in taste, and has been documented to confer bitterness in green coffee beans (Baeza et al., 2016).

The present study also detected methanol as a potential biomarker, separating the Comm from the NComm cultivars. In a study by Fall and Benson, (1996), it was documented that methanol is a natural metabolism product, emitted from plant leaves. Their study showed high concentrations of methanol in the forest air, similar to results reported in a separate study by Fehsenfeld et al., (1992), further substantiating the likelihood that the forest plants were producing this compound. Employing GC analysis or direct enzymatic analysis of gas-phase methanol, significant methanol emissions were observed from the leaves of forest

plants using leaf and branch enclosure approaches, typically ranging between 0.0003 - 0.017 mg methanol per g dry weight (MacDonald and Fall, 1993), with young leaves emitting up to 0.04 mg methanol per g dry weight (Nemecek-Marshall et al., 1995). It has been postulated that the source of methanol in the leaves is cell wall pectin demethylation (O'neill et al., 1990). Methanol has been shown to possess a sweet taste and is used in artificial sweeteners (Chattopadhyay et al., 2014). The present study found the concentrations of methanol in the Comm cultivars to be double that of the NCom, at 0.2 and 0.1 mg/g, respectively.

Glycerol is another metabolite that was detected on the GC-MS platform. Literature shows that glycerol enhances the aroma and sweetness levels in wines. In a study investigating the effects of glycerol in red and white wines, it was documented that a glycerol concentration of 10 g/L was sufficient to enhance the aroma, and suppress the bitterness, resulting in these wines being reported as sweet-tasting even when glucose and fructose levels were below the detection threshold reported in other studies (Jones et al., 2008; Sokolowsky and Fischer, 2012). In earlier, similar studies, the glycerol sweetness thresholds for dry white wine were reported as 5.2 mg/g (Noble and Bursick, 1984) and 9 mg/g (Hinreiner et al., 1955). This study found no significant difference in the concentrations of glycerol in the Comm and NComm cultivars.

Since several metabolites described above relate to sweetness, the summation of the metabolites; arabinose, arabitol, glycerol, malic acid, phloroglucinol, psicose, ribitol, sucrose, and xylonic acid for the GC-MS, and glucose, methanol and sucrose for the ^{1}H-NMR, in each of the 310 cultivars was calculated. This sum of sweeteners was labelled total sweeteners. Interestingly, the total sweeteners were significantly higher ($P<0.05$) in Comm cultivars. We report here for the first time in the leaves of black tea cultivars that several metabolites, related to sweetness, which are higher in the Comm than the NComm cultivars. Sweetness may have contributed to these cultivars being selected for commercial production since the 1950s.

As documented below, several metabolites were also detected that are responsible for other taste qualities i.e. bitterness and umami. Plants are rich in bitter-tasting metabolites, which serve to deter herbivores. To reduce bitterness or off-taste, plants produce sweet, acidic, or strong fruity flavoured molecules, which mask the bad tastes (Ley, 2008; Tripathi et al., 2011). The present study shows that the Comm cultivars have a higher total sweeteners concentration of 9.3 mg/g, which is significantly higher ($P<0.05$) from the total sweeteners concentration in the NComm of 8.7 mg/g dry weight. Furthermore, the total amino acid

concentration, calculated by adding the amino acids alanine, isoleucine, leucine, theanine, and valine, was higher in the Comm than the NCom, at 21.8 and 20.9 mg/g dry weight. These compounds could, therefore, mask the bitterness, resulting from e.g. caffeine and chlorogenic acid. As mentioned in the foregoing, amino acids are responsible for the aroma of tea, therefore, the higher the total amino acids, the more aromatic the tea. The total amino acid and total sweeteners concentrations being higher in the Comm cultivars result in the teas produced from these cultivars having a better taste.

Linolenic acid was another metabolite detected by the GC-MS, which was higher but not statistically significant, in the Comm cultivars. Linolenic acid has been documented as being responsible for the bitter taste observed in soybean lecithins, and linseed oil obtained from different varieties (Stephan and Steinhart, 2000; Brühl et al., 2008). A study by Toumi et al., (2008), revealed linolenic acid to be the most abundant fatty acid in grapevine leaves, and it was higher in the leaves of DT cultivars, which corroborates our findings. The total lipid membrane composition in plants increases during drought, reducing the amount of water lost by the plant. It is therefore no surprise that DT cultivars have higher linolenic acid content, as it serves as a mechanism for coping with drought stress (Welti et al., 2002). Threonic acid is a by-product of ascorbate catabolism, whose production is induced, and regulated by stresses such as light and drought; this metabolite confers protection to the plant (Renault et al., 2017). It has been documented to be involved in stomatal closure during drought stress (Valpuesta and Botella, 2004). The levels of threonic acid found in this study were significantly higher ($P<0.05$) in the Comm cultivars than the NComm cultivars. This is expected as threonic acid has been reported to confer drought tolerance properties in plants, and as such it can be postulated that this metabolite contributes to the drought tolerance properties observed in the Comm cultivars.

Correlations have been documented in the literature between the umami taste found in green tea and theanine. Theanine concentrations have been reported to range between 10-50 mg/g dry weight (Vuong et al., 2011) in *C. sinensis* and the higher the concentration, the more the umami taste. The average theanine levels detected in this study were higher, but insignificant, in the Comm (8.6 mg/g) compared to the NComm cultivars (8.0 mg/g) dry weight. Gallic acid was detected to have slightly higher concentrations in the Comm cultivars (1.1 mg/g) than the NComm (1.0 mg/g). In a study by Kaneko et al. (2006), the umami taste intensity of green tea without any additives was evaluated and was scored at intensity of 1.5 out of five. However, following the addition of 5.4 mg/g of natural gallic acid, the intensity of the umami

taste increased to a score of 2.4. This higher umami score indicates the significance of gallic acid in the taste of tea. According to literature, fresh green tea leaves contain trace amounts of gallic acid, which then increase during the manufacturing process of black tea as a result of galloyl ester hydrolysis from the catechins and theaflavin gallates. The high levels of gallic acid in some cultivars have been attributed to correspondingly high levels of gallated catechins, which result in the generation and consequent degradation of the theaflavins. It has been documented that EGCg and ECg are principal taste metabolites in tea, which are responsible for tea astringency, while caffeine is responsible for bitterness (Xu et al., 2018). The gallated catechins ECg and EGCg significantly contribute to the generation of theaflavins in black tea. As such, high concentrations of ECg and EGCg may be markers for high-quality black teas. It has been reported that the ratio of di-hydroxyl flavan-3-ols to tri-hydroxyl flavan-3-ols impacted the quality of black tea; high quantities of simple catechins such as catechin, EC and ECg compared to the gallo-catechins EGC and EGCg, results in higher amounts of theaflavins (Kwach et al*.,* 2016). According to Hilton and Palmer-Jones (1973), the concentration of EGC in the fresh tea shoots strongly correlates with theaflavins amounts, thus influencing black tea pricing (Owuor and Obanda, 2007). The higher EGC found in the Comm cultivars agrees with Owuor and Obanda's findings (Owuor and Obanda, 2007).

From the results obtained in the present study, the Comm cultivars had a significantly ($P<0.05$) higher total catechins content of 121.58 mg/g compared to the 114.96 mg/g dry weight of the NComm cultivars. The total catechins were calculated by adding CAT, EC, ECg, EGC, and EGCg. The catechins are the substrates from which theaflavins and thearubigins are produced during the manufacture of black tea. Theaflavins are responsible for the astringency, another important trait in the quality determination of tea. The theaflavin digallate, is approximately 6.4 times more astringent than theaflavin, while also being 2.88 times more astringent than both theaflavin-3-monogallate and theaflavin-3'-monogallate (Obanda et al., 2001). According to Lin et al*.,* (1996), a 200 mL cup of green tea contains 305 mg EGCg, 145 mg EGC, 70 mg ECg, 28 mg EC, and 8 mg CAT. It can, therefore, be concluded that the total catechins may have contributed to the selection of the Comm cultivars since the 1950s.

5. Recommendation and conclusion

Further targeted metabolomics studies are warranted on these metabolites reported in Table 1. This study successfully identified metabolites expressed differently in the Comm and NComm cultivars. The metabolites can be used for selection of high quality black and drought tolerant tea cultivars.

6. Acknowledgements

The authors acknowledge the financial support to conduct this research, and study grants for CN from James Finlay (Kenya) Ltd., George Williamson (Kenya) Ltd., Sotik Tea Company (Kenya) Ltd., Mcleod Russell (Uganda) Ltd., the TRI of Kenya, and *Southern African Biochemistry and Informatics for Natural Products (*SABINA). The *C. sinensis* cultivars used in this study were provided by the TRI of Kenya. Supplementary funding was provided by the Technology and Human Resources for Industry Programme (THRIP), an initiative of the Department of Trade and Industries of South Africa (dti), the National Research Foundation (NRF) of South Africa, and the University of Pretoria South Africa.

7. Author contributions

ZA, RM and SK were involved with the experimental design of the research. RK and CN were responsible for plant material collection. CN, SM, and ZL conducted the experiments. MvR performed statistical analysis. CN wrote the manuscript, which was revised by MvR, RK, RM, SK, SM, ZL, and ZA. The manuscript was reviewed and approved by all the authors.

8. Compliance with ethical standards

7.1 Conflict of interest

The authors assert that they have no conflicts of interest.

9. References

Ackermann, J., Fischer, M., Amado, R., 1992. Changes in sugars, acids, and amino acids during ripening and storage of apples (cv. Glockenapfel). J. Agric. Food Chem. 40(7), 1131-1134.

Adkins, N. L., Hall, J. A., Georgel, P. T., 2007. The use of quantitative agarose gel electrophoresis for rapid analysis of the integrity of protein–DNA complexes. J. Biochem. Biophys. Met. 70(5), 721-726.

Archana, I. and Vijayalakshmi, K., 2018. Antioxidant potential of phloroglucinol; an in-vitro approach. Int. J. Pharm. Sci. Res. 9(7), 2947-2951.

Baeza, G., Sarriá, B., Bravo, L., Mateos, R., 2016. Exhaustive Qualitative LC-DAD-MS n Analysis of Arabica Green Coffee Beans: Cinnamoyl-glycosides and Cinnamoylshikimic Acids as New Polyphenols in Green Coffee. J. Agric. Food Chem. 64(51), 9663-9674.

Bandurski, R. S., and Schulze, A., 1977. Concentration of indole-3-acetic acid and its derivatives in plants. Plant Phys. 60(2), 211-213.

Borsani, J., Budde, C. O., Porrini, L., Lauxmann, M. A., Lombardo, V. A., Murray, R., Andreo, C. S., Drincovich, M. F., Lara, M. V., 2009. Carbon metabolism of peach fruit after harvest, changes in enzymes involved in organic acid and sugar level modifications. J. Exp. Bot. 60(6), 1823-1837.

Brühl, L., Matthäus, B., Scheipers, A., Hofmann, T., 2008. Bitter off-taste in stored cold-pressed linseed oil obtained from different varieties. Eur. J. Lip. Sci. Technol. 110(7), 625-631.

Buffo, R. A., and Cardelli-Freire, C., 2004. Coffee flavour, an overview. Flavour. Frag. J. 19(2), 99-104.

Cabrera, C., Artacho, R., Giménez, R., 2006. Beneficial effects of green tea—a review. JACN. 25(2), 79-99.

Chattopadhyay, S., Raychaudhuri, U., Chakraborty, R., 2014. Artificial sweeteners–a review. JFST. 51(4), 611-621.

Chaturvedula, V. S. P. and Prakash, I., 2011. The aroma, taste, color and bioactive constituents of tea. J. Med. Plant. Res. 5(11), 2110-2124.

Chin, J. M., Merves, M. L., Goldberger, B. A., Sampson-Cone, A., Cone, E. J., 2008. Caffeine content of brewed teas. J. Anal. Toxicol. 32(8), 702-704.

Chugh, K., 2013.Measuring phenotypic and genetic variances and narrow sense heritability in three populations of annual ryegrass (Lolium multiflorum Lam.).

Clifford, M. N., 2000. Chlorogenic acids and other cinnamates–nature, occurrence, dietary burden, absorption and metabolism. J. Sci. Food Agric. 80(7), 1033-1043.

Cortell, J. M., Sivertsen, H. K., Kennedy, J. A., Heymann, H., 2008. Influence of vine vigor on Pinot noir fruit composition, wine chemical analysis, and wine sensory attributes. AJEV. 59(1), 1-10.

Dumas, M.-E., Maibaum, E. C., Teague, C., Ueshima, H., Zhou, B., Lindon, J. C., Nicholson, J. K., Stamler, J., Elliott, P., Chan, Q., 2006. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research, the INTERMAP Study. Anal. Chem. 78(7), 2199-2208.

Dunn, W. B., Goodacre, R., Neyses, L., Mamas, M., 2011. Integration of metabolomics in heart disease and diabetes research, current achievements and future outlook. Bioanal. 3(19), 2205-2222.

Dutta, R., Stein, A., Bhagat, R., 2011. Integrating satellite images and spectroscopy to measuring green and black tea quality. Food Chem. 127(2), 866-874.

Ebbels, T. M., Lindon, J. C., Coen, M. 2011. Processing and modeling of nuclear magnetic resonance (NMR) metabolic profiles. In Metabolic Profiling, 365-388, Springer.

Ellinger, J. J., Chylla, R. A., Ulrich, E. L., Markley, J. L., 2013. Databases and software for NMR-based metabolomics. Curr. Metabolomics 1(1), 28-40.

Ellis, R. and Nyirenda, H., 1995. A successful plant improvement programme on tea (Camellia sinensis). Exp. Agric. 31(3), 307-323.

Ellis, S.M. and Steyn, H.S., 2003. Practical significance (effect sizes) versus or in combination with statistical significance (p-values), Management Dynamics 12(4), 51-53.

Fall, R., and Benson, A. A., 1996. Leaf methanol—the simplest natural product from plants. Trends Plant Sci. 1(9), 296-301.

FAO (Food and Agricultural Organisation)., 2015. Retrieved from http, //www.fao.org.

Fehsenfeld, F., Calvert, J., Fall, R., Goldan, P., Guenther, A. B., Hewitt, C. N., ..., Zimmerman, P., 1992. Emissions of volatile organic compounds from vegetation and the implications for atmospheric chemistry. Global Biogeochem. Cy. 6(4), 389-430.

Goddijn, O. J., and van Dun, K., 1999. Trehalose metabolism in plants. Trends Plant Sci. 4(8), 315-319.

Gramza, A., Khokhar, S., Yoko, S., Gliszczynska-Swiglo, A., Hes, M., Korczak, J., 2006. Antioxidant activity of tea extracts in lipids and correlation with polyphenol content. Eur. J. Lip. Sci. Technol. 108(4), 351-362.

Group, C. P. B., Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L., Chen, S.-L., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. PNAS. 108(49), 19641-19646.

Habibi, Y., Mahrouz, M., Marais, M. F., Vignon, M. R., 2004. An arabinogalactan from the skin of Opuntia ficus-indica prickly pear fruits. Carb. Res. 339(6), 1201-1205.

Hamanishi, E. T., Barchet, G. L., Dauwe, R., Mansfield, S. D., Campbell, M. M., 2015. Poplar trees reconfigure the transcriptome and metabolome in response to drought in a genotype-and time-of-day-dependent manner. BMC Genomics 16(1), 329.

Hilton, P. J., and Palmer-Jones, R., 1973. Relationship between the flavanol composition of fresh tea shoots and the theaflavin content of manufactured tea. J. Sci. Food. Agric. 24(7), 813-818.

Hinreiner, E., Filipello, F., Berg, H. W., Webb, A. D., 1955. Evaluation of thresholds and minimum difference concentrations for various constituents of 4 wines. Detectable differences in wine. Food Technol. 9(10), 489-490.

Hoekstra, F. A., Golovina, E. A., Buitink, J., 2001. Mechanisms of plant desiccation tolerance. Trends Plant Sci. 6(9), 431-438.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65-70.

Horemans, N., Foyer, C. H., Potters, G., Asard, H., 2000. Ascorbate function and associated transport systems in plants. Plant Phys. Biochem. 38(7-8), 531-540.

Iiyama, S., Ezaki, S., Toko, K., Matsuno, T., Yamafuji, K., 1995. Study of astringency and pungency with multichannel taste sensor made of lipid membranes. Sensor. Actuat. B-Chem. 24(1-3), 75-79.

Imamura, K., Tsuchama, Y., Tsunekawa, H., Okamura, K., Okamoto, R., Harada, K., 1995. Removal of chlorogenic acid from coffee extracts. Jpn. Kokai JP 07322823.

Inglett, G. E., 1971. Intense sweetness of natural origin. In: G.G. Birch, L.F. Green, and C.B. Coulson (eds.). Sweetness and sweeteners. Elsevier Appl. Sci., London.

International Tea Committee (ITC). Annual Bulletin of Statistics, 2019.

Jaiswal, R., Sovdat, T., Vivan, F., Kuhnert, N., 2010. Profiling and characterization by LC-MS n of the chlorogenic acids and hydroxycinnamoylshikimate esters in mate (Ilex paraguariensis). J. Agric. Food Chem. 58(9), 5471-5484.

Jones, P. R., Gawel, R., Francis, I. L., Waters, E. J., 2008. The influence of interactions between major white wine components on the aroma, flavour and texture of model white wine. Food Qual. Prefer. 19(6), 596-607.

Kaneko, S., Kumazawa, K., Masuda, H., Henze, A., Hofmann, T., 2006. Molecular and sensory studies on the umami taste of Japanese green tea. J. Agric. Food Chem. 54(7), 2688-2694.

Kang, J., Choi, M.-Y., Kang, S., Kwon, H. N., Wen, H., Lee, C. H., Park, M., Wiklund, S., Kim, H. J., Kwon, S. W., 2008. Application of a 1H nuclear magnetic resonance (NMR) metabolomics approach combined with orthogonal projections to latent structure-discriminant analysis as an efficient tool for discriminating between Korean and Chinese herbal medicines. J. Agric. Food Chem. 56(24), 11589-11595.

Kaplan, F. and Guy, C. L., 2004. β-Amylase induction and the protective role of maltose during temperature shock. Plant Phys. 135(3), 1674-1684.

Kenya National Bureau of Statistics, 2018.Kenya facts and figures 2018. Kenya.

Kerchev, P. I., Fenton, B., Foyer, C. H., Hancock, R. D., 2012. Plant responses to insect herbivory, interactions between photosynthesis, reactive oxygen species and hormonal signalling pathways. Plant Cell Environ. 35(2), 441-453.

Khan, N. and Mukhtar, H., 2007. Tea polyphenols for health promotion. Life Sci. 81(7), 519-533.

Kobayashi-Hattori, K., Mogi, A., Matsumoto, Y., Takita, T., 2005. Effect of caffeine on the body fat and lipid metabolism of rats fed on a high-fat diet. Biosci Biotech Bioch 69(11), 2219-2223.

Koech, R. K., Malebe, P. M., Nyarukowa, C., Mose, R., Kamunya, S. M., Apostolides, Z., 2018. Identification of novel QTL for black tea quality traits and drought tolerance in tea plants (*Camellia sinensis*). Tree Genet. Genomes 14(1), 9.

Kowalsick, A., Kfoury, N., Robbat Jr, A., Ahmed, S., Orians, C., Griffin, T., Cash, S. B., Stepp, J. R., 2014. Metabolite profiling of Camellia sinensis by automated sequential, multidimensional gas chromatography/mass spectrometry reveals strong monsoon effects on tea constituents. J. Chrom. A. 1370, 230-239.

Kumar, V., Rani, A., Goyal, L., Pratap, D., Billore, S. D., Chauhan, G. S., 2011. Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavones contents. Int. J. Food Prop. 14(5), 1142-1151.

Kwach, B. O., Owuor, P. O., Kamau, D. M., Msomba, S. W., Uwimana, M. A., 2016. Variations in the precursors of plain black tea quality parameters due to location of production and nitrogen fertilizer rates in Eastern African clonal tea leaves. Exp. Agric. 52(2), 266-278

Lavarack, B., Griffin, G., Rodman, D., 2002. The acid hydrolysis of sugarcane bagasse hemicellulose to produce xylose, arabinose, glucose and other products. Biomass Bioenerg. 23(5), 367-380.

Le Gall, G., Colquhoun, I. J., Defernez, M., 2004. Metabolite profiling using 1H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). J. Agric. Food Chem. 52(4), 692-700.

Ley, J. P., 2008. Masking bitter taste by molecules. Chemosens. Percept. 1(1), 58-77.

Lin, Y. L., Juan, I. M., Chen, Y. L., Liang, Y. C., Lin, J. K., 1996. Composition of polyphenols in fresh tea leaves and associations of their oxygen-radical-absorbing capacity with antiproliferative actions in fibroblast cells. J. Agric. Food Chem. 44(6), 1387-1394.

Lorist, M. M., and Tops, M., 2003. Caffeine, fatigue, and cognition. Brain Cogn. 53(1), 82-94.

Ludwig, C. and Viant, M. R., 2010. Two-dimensional J-resolved NMR spectroscopy, review of a key methodology in the metabolomics toolbox. Phytochemical Analysis, Int. J. Plant Chem. Biochem. Tech. 21(1), 22-32.

Ma, B., Chen, J., Zheng, H., Fang, T., Ogutu, C., Li, S., Han, Y., Wu, B., 2015. Comparative assessment of sugar and malic acid composition in cultivated and wild apples. Food Chem. 172, 86-91.

MacDonald, R. C., and Fall, R., 1993. Detection of substantial emissions of methanol from plants to the atmosphere. A. Gen. Top. 27(11), 1709-1713.

Matanjun, P., Mohamed, S., Mustapha, N. M., Muhammad, K., Ming, C. H., 2008. Antioxidant activities and phenolics content of eight species of seaweeds from north Borneo. J. Appl. Phycol. 20(4), 367-373.

Marks, S. C., Mullen, W., Crozier, A., 2007. Flavonoid and chlorogenic acid profiles of English cider apples. J. Sci. Food Agric. 87(4), 719-728.

Mazzafera, P., and Silvarolla, M. B., 2010. Caffeine content variation in single green Arabica coffee seeds. Seed Sci. Res. 20(3), 163-167.

Melgarejo, P., Salazar, D. M., Artes, F., 2000. Organic acids and sugars composition of harvested pomegranate fruits. Eur. Food Res. Technol. 211(3), 185-190.

Min, L. U., Huaming, A. N., Daoping, W. A. N. G., 2017. Characterization of amino acid composition in fruits of three Rosa roxburghii genotypes. Hortic. Plant J. 3(6), 232-236.

Mu, W., Zhang, W., Feng, Y., Jiang, B., Zhou, L., 2012. Recent advances on applications and biotechnological production of D-psicose. Appl. Micro. Biotech. 94(6), 1461-1467.

Nabors, L. O., and Gelardi, R., 2001. Alternative sweeteners, An overview. Food science and technology-New York-Marcel Dekker 1-12.

Nara, K., Kato, Y., and Motomura, Y., 2001. Involvement of terminal-arabinose and-galactose pectic compounds in mealiness of apple fruit during storage. Postharvest Biol. Tec. 22(2), 141-150.

Naveed, M., Brown, L. K., Raffan, A. C., George, T. S., Bengough, A. G., Roose, T., ..., Hallett, P. D., 2017. Plant exudates may stabilize or weaken soil depending on species, origin and time. Eur. J. Soil Sci. 68(6), 806-816.

Nemecek-Marshall, M., MacDonald, R. C., Franzen, J. J., Wojciechowski, C. L., Fall, R., 1995. Methanol emission from leaves (enzymatic detection of gas-phase methanol and relation of methanol fluxes to stomatal conductance and leaf development). Plant Phys. 108(4), 1359-1368.

Nitin Seetohul, L., Islam, M., O'Hare, W. T., Ali, Z., 2006. Discrimination of teas based on total luminescence spectroscopy and pattern recognition. J. Sci. Food Agric. 86(13), 2092-2098.

Noble, A. C., and Bursick, G. F., 1984. The contribution of glycerol to perceived viscosity and sweetness in white wine. AJEV. 35(2), 110-112.

Nyarukowa, C., Koech, R., Loots, T., Apostolides, Z., 2016. SWAPDT, A method for Short-time Withering Assessment of Probability for Drought Tolerance in Camellia sinensis validated by targeted metabolomics. J. Plant Phys. 198, 39-48.

Nyarukowa, C. T., Koech, K. R., Loots, T., Hageman, J., Apostolides, Z., 2018. Prioritising the replanting schedule of seedling tea fields on tea estates for drought susceptibility measured by the SWAPDT method in the absence of historical in-filling records. J. Agric. Sci. 10(7), 26-34.

Obanda, M., Owuor, P. O., Mang'oka, R., 2001. Changes in the chemical and sensory quality parameters of black tea due to variations of fermentation time and temperature. Food Chem. 75(4), 395-404.

O'Neill, M., Albersheim, P., Darvill, A. 1990. The pectic polysaccharides of primary cell walls. P. M. Dey (Ed.), Methods in Plant Biochemistry, 2, Academic Press, London, 415-441.

Oshima, H., Kimura, I., Izumori, K., 2006. Psicose contents in various food products and its origin. Food Sci. Tech. Res. 12(2), 137-143.

Owuor, P. O., and Obanda, M., 2007. The use of green tea (Camellia sinensis (L)) leaf flavan-3-ols composition in predicting plain black tea quality potential. Food Chem. 100(3), 873–884.

Pongsuwan, W., Fukusaki, E., Bamba, T., Yonetani, T., Yamahara, T., Kobayashi, A., 2007. Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. J. Agric. Food Chem. 55(2), 231-236.

Preedy, V. R., 2013. Tea in health and disease prevention. Academic Press, London.

Qin, Z., Pang, X., Chen, D., Cheng, H., Hu, X., Wu, J., 2013. Evaluation of Chinese tea by the electronic nose and gas chromatography–mass spectrometry, Correlation with sensory properties and classification according to grade level. Food Res. Int. 53(2), 864-874.

Rawat, R., Gulati, A., Babu, G. K., Acharya, R., Kaul, V. K., Singh, B., 2007. Characterization of volatile components of Kangra orthodox black tea by gas chromatography-mass spectrometry. Food Chem. 105(1), 229-235.

Renault, H., Alber, A., Horst, N. A., Lopes, A. B., Fich, E. A., Kriegshauser, L., Pineau, E., 2017. A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nature Comm. 8, 14713.

Rodrigues, C. I., Marta, L., Maia, R., Miranda, M., Ribeirinho, M., Máguas, C., 2007. Application of solid-phase extraction to brewed coffee caffeine and organic acid determination by UV/HPLC. J. Food Comp. Anal. 20(5), 440-448.

Roser, D. J., Melick, D. R., Ling, H. U., Seppelt, R. D. (1992). Polyol and sugar content of terrestrial plants from continental Antarctica. Antarct. Sci. 4(4), 413-420.

Saccenti, E., Hoefsloot, H. C., Smilde, A. K., Westerhuis, J. A., Hendriks, M. M., 2014. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics, 10(3), 361-374.

Sanderson, G. W., Grahamm, H. N., 1973. Formation of black tea aroma. J. Agric. Food Chem. 21(4), 576-585.

Sanhueza, E., and Andreae, M. O., 1991. Emission of formic and acetic acids from tropical savanna soils. Geophys. Res. Lett. 18(9), 1707-1710.

Schuh, C. and Schieberle, P., 2006. Characterization of the key aroma compounds in the beverage prepared from Darjeeling black tea, quantitative differences between tea leaves and infusion. J. Agric. Food Chem. 54(3), 916-924.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry, communicating confidence. Environ. Sci. Technol. 48(4), 2097-2098.

Sokolowsky, M., and Fischer, U., 2012. Evaluation of bitterness in white wine applying descriptive analysis, time-intensity analysis, and temporal dominance of sensations analysis. Anal. Chim. Acta. 7(32), 46-52.

Stephan, A. and Steinhart, H., 2000. Bitter taste of unsaturated free fatty acids in emulsions, contribution to the off-flavour of soybean lecithins. Eur. Food Res. Technol. 212(1), 17-25.

Szejtli, J., and Szente, L., 2005. Elimination of bitter, disgusting tastes of drugs and foods by cyclodextrins. Eur. J. Pharma. Biopharma. 61(3), 115-125.

Tan, F., Tan, C., Zhao, A., Li, M., 2011. Simultaneous determination of free amino acid content in tea infusions by using high-performance liquid chromatography with fluorescence detection coupled with alternating penalty trilinear decomposition algorithm. J. Agric. Food Chem. 59(20), 10839-10847.

Taylor, N. W., 1928. A physico-chemical theory of sweet and bitter taste excitation based on the properties of the plasma membrane. Protoplasma 4(1), 1-17.

Toumi, I., Gargouri, M., Nouairi, I., Moschou, P. N., Salem-Fnayou, A. B., Mliki., ..., Ghorbel, A., 2008. Water stress induced changes in the leaf lipid composition of four grapevine genotypes with different drought tolerance. Biologia Plantarum. 52(1), 161-164.

Tripathi, A., Parmar, D., Patel, U., Patel, G., Daslaniya, D., Bhimani, B., 2011. Taste masking, a novel approach for bitter and obnoxious drugs. JPSBR 1(3), 36-142.

Unno, K., Tanida, N., Ishii, N., Yamamoto, H., Iguchi, K., Hoshino, M., ..., Yamada, H., 2013. Anti-stress effect of theanine on students during pharmacy practice, positive correlation among salivary α-amylase activity, trait anxiety and subjective stress. Pharmacol. Biochem. Behav. 111, 128-135.

Valpuesta, V., and Botella, M. A., 2004. Biosynthesis of L-ascorbic acid in plants: new pathways for an old antioxidant. Trends Plant Sci. 9(12), 573-577.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., van der Werf, M. J., 2006. Centering, scaling, and transformations, improving the biological information content of metabolomics data. BMC genomics 7(1), 142.

Vencl, F. V., and Morton, T. C., 1998. The shield defense of the sumac flea beetle, Blepharida rhois (Chrysomelidae: Alticinae). Chemoecology. 8(1), 25-32.

Vuong, Q. V., Bowyer, M. C., Roach, P. D., 2011. L-Theanine, properties, synthesis and isolation from tea. J. Sci. Food Agric. 91(11), 1931-1939.

Warrack, B. M., Hnatyshyn, S., Ott, K. H., Reily, M. D., Sanders, M., Zhang, H., Drexler, D. M., 2009. Normalization strategies for metabonomic analysis of urine samples. J. Chromatogr. B. 877(5-6), 547-552.

Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H. E., ..., Wang, X., 2002. Profiling membrane lipids in plant stress responses role of phospholipase Dα in freezing-induced lipid changes in Arabidopsis. J. Bio. Chem. 277(35), 31994-32002.

Worley, B., Halouska, S., Powers, R., 2013. Utilities for quantifying separation in PCA/PLS-DA scores plots. Anal. Biochem. 433(2), 102-104.

Xu, Y.-Q., Zhang, Y.-N., Chen, J.-X., Wang, F., Du, Q.-Z., Yin, J.-F., 2018. Quantitative analyses of the bitterness and astringency of catechins from green tea. Food Chem. 258, 16-24.

Yan, S. H., 2007. NIR evaluation of the quality of tea and its market price. Spectr. Eur. 19(2), 16-19.

Yoshida, T. and Sakamoto, T., 2009. Water-stress induced trehalose accumulation and control of trehalase in the cyanobacterium Nostoc punctiforme IAM M-15. JGAM. 55(2), 135-145.

Yu, H., Wang, J., Yao, C., Zhang, H., Yu, Y., 2008. Quality grade identification of green tea using E-nose by CA and ANN. LWT-Food Sci. Technol. 41(7), 1268-1273.

Zentella, R., Mascorro-Gallardo, J. O., Van Dijck, P., Folch-Mallol, J., Bonini, B., Van Vaeck, C., ..., Iturriaga, G., 1999. A Selaginella lepidophylla trehalose-6-phosphate synthase complements growth and stress-tolerance defects in a yeast tps1 mutant. Plant Phys. 119(4), 1473-1482.

Zhang, J., Wang, X., Yu, O., Tang, J., Gu, X., Wan, X., Fang, C., 2010. Metabolic profiling of strawberry (Fragaria × ananassa Duch.) during fruit development and maturation. J. Exp. Bot. 62(3), 1103-1118.

Zhou, J., Ho, C. T., Long, P., Meng, Q., Zhang, L., Wan, X., 2019. Preventive Efficiency of Green Tea and Its Components on Nonalcoholic Fatty Liver Disease. J. Agric. Food Chem. 67(19), 5306-5317.