Genomic prediction and genome wide association mapping of quality traits in tea (*Camellia sinensis* (L.) O. Kuntze)

LUBANGA NELSON MANDELA

Thesis submitted to The University of Nottingham for the degree of

Doctor of Philosophy

October 2019

Dedication

My daughter, Olivia

Acknowledgements

I would like to express my sincere appreciation to my academic supervisors **Prof. Festo Massawe**, **Dr. Sean Mayes and Dr. Jacquie De Silva** for their outstanding mentorship, great inspiration, support and wonderful encouragement throughout my Ph.D. I am privileged to be mentored by these three supervisors. A special thanks to **Dr. Ho Wai Kuan** and **Dr. Maysoun Mustapha** at Crops for the future in Malaysia who provided a very supportive environment. I am also thankful to **Dr. Katie Mayes** and **Chris De Silva** for their hospitality while in United Kingdom.

I am also grateful to **Kenneth Odire**, **Gabriel Tuwei**, **Faith Maluki and Jack Ronge** who provided huge support for my research at Unilever Tea Kenya (UTK). I would also like to express my gratitude to all the staff at the plant improvement center (PIC) and the laboratory staff at Research & Development, UTK for all the support. Likewise, I sincerely thank **Sally Wallace** at R&D Colworth for helping with the sample preparation for NMR analysis.

I am indebted to Unilever R&D Colworth and University of Nottingham, Malaysia campus for funding my Ph.D. and research. The scholarship was instrumental in shaping my research and professional career, and I will forever be grateful. I am also grateful to ADnid laboratories in France for DNA extraction, Cornell University's Institute of Genomic Diversity Facility for genotyping (GBS), Nature Source Improved Plants (NSIP) for SNP calling and Rothamsted Research for NMR analysis.

No words can express my deep appreciations to my parents **Jane** and **Henry**, who gave me all the blessings and were always there for me. I am also very grateful to my siblings **Christine, Dickens, Virginia** and **Gilbert**. My late father in law **Bramwel Ekwang** was so passionate about my Ph.D. and I had promised him I would finish before this year ends. I am also grateful for the support and encouragement of my mother in law, **Betty Andeyo**. Finally, I owe very special endless appreciations to my dear wife **Carren Mercy Ekwang** for her continuous support and encouragement that gave me the strength to continue this work and for considering my success her own.

Abstract

Conventional tea (*Camellia sinensis* (L.) O. Kuntze) breeding consisting of recurrent cycles of crossing, field evaluation and phenotypic selection is the main breeding technique for tea. However, it is a time-consuming process, that result in slow genetic gain. In order to accelerate tea breeding, the use of modern breeding methods is required. In this regard, genomic selection (GS) and genome wide association studies (GWAS) has been considered most promising for genetic improvement of complex traits. The main aim of our study was to investigate the applicability of genomic selection (GS) and genome wide association studies (GWAS) in tea breeding. A training population consisting of 103 tea genotypes located at two sites were genotyped using genotyping by sequencing (GBS). Twelve biochemical traits known to influence tea quality were evaluated using Nuclear Magnetic resonance (NMR) spectroscopy. Hierarchical cluster and principle component analyses distinguished the 103 genotypes based on their biochemical properties. Additionally, specific biochemical compounds correlated with sensory properties; mouthfeel and taste correlated with ECG and EGCG, respectively. This implies that, biochemical compounds could be used for selecting high quality teas objectively at the seedlings stage in the nursery, hence saving the time. We also concluded that an optimized miniature process could be used for manufacturing different tea varieties into black tea, however technologies that could optimally control withering and fermentation steps for the different tea varieties developed in a breeding programme could be explored further. Using GWAS, we identified 64 significant SNP markers and candidate genes associated with the biochemical traits. The potential candidate genes identified included transferases, cytochrome P450 704C1-like proteins, E3 ubiquitin protein ligases, ATP-dependent zinc metalloprotease and exopolygalacturonases. The candidate genes and the associated SNPs provide valuable resources for future studies to breed high quality tea varieties and to understand the genetic basis of tea quality at a chemical level, to complement the current sensory method of tea-tasting. The identified SNP markers could be further fine mapped to evaluate their potential involvement in tea quality. Among the 2779 sequence tags, only 929 SNPs were mapped to each of the two published draft genomes. In addition, 311 sequences had blast hits while 217 sequences were

annotated and were assigned to biological processes, cellular component and molecular functions. We also compared the prediction accuracies of 5 GS models using a 5-fold cross validation approach. However, the performance of all the GS models were almost the same, with RRBLUP, BayesLASSO and BayesA performing slightly better than BayesB and BayesCπ. Traits with high GS accuracies were Epigallocatechin gallate (ECGG), Theanine, Epicatechin (EC), Epicatechin gallate (EGC) and theobromine, while those with low prediction accuracies were Gallocatechin (GC), catechin and Gallic acid (GA). We conclude that implementing GWAS and GS in tea breeding would help to improve the prediction accuracies and benefit from rapid genetic gains from selection of high-quality teas.

Table of content

List of tables

List of figures

List of abbreviations and symbols

ANOVA:	Analysis of variance
Bayes A:	Bayesian A
Bayes B:	Bayesian B
Bayes Cp:	Bayesian Cp
Bayes Lasso:	Bayesian least absolute shrinkage and selection operator
BLAST:	Basic local alignment tool
BLUE:	Best linear unbiased estimates
BLUP:	Best linear unbiased prediction
bp:	base pair
C: (+)	Catechin
CFT:	Clonal field trial
CSS:	*Camellia sinensis* var**.** *sinensis*
CSA:	Camellia *sinensis* var. *asamica*
CTC:	Cut, tear and curl
DNA:	Deoxyribonucleic acid
EBVs:	Estimated breeding values
EC: (-)	Epicatechin
ECG: (-)	Epicatechin gallate
EGC: (-)	Epigallocatechin
EGCG: (-)	Epigallocatechin gallate
GA:	Gallic acid
GBLUP:	Genomic best linear unbiased prediction;
GBS:	Genotyping-by-sequencing

GC:	Gallocatechin
GCG:	Gallocatechin gallate
GEBV:	Genomic estimated breeding values
GS:	Genomic selection
GWAS:	Genome-wide association study
HPT:	Hand pollination trial
LD:	Linkage disequilibrium
MAS:	Marker assisted selection
MCT:	Mini clonal trial
NMR:	Nuclear magnetic resonance
PCA:	Principal component analysis
r^2:	Coefficient of linkage disequilibrium
REML:	Restricted maximum likelihood
RR-BLUP:	Ridge regression-best linear unbiased prediction
SNP:	Single nucleotide polymorphism
TASSEL:	Trait Analysis by aSSociation, Evolution and Linkage
TP:	Training population
UTK:	Unilever Tea, Kenya
σ^2a:	Genotypic variance
σ^2e:	Error variance
σ^2sg:	Variance due to genotype by environment interaction

Thesis structure

The write-up of the thesis follows 'thesis by published work' format. A total of five papers (draft manuscripts) are included in the main body of thesis (Table 1). The thesis is divided into seven main sections, namely;

1. **General introduction:** This section contains background information, problem statements, aims and objective(s) of the study.

2. **Literature review:** This section has detailed information on tea, tea quality, molecular markers and the breeding techniques applied in tea.

3. **Thematic area one:** This section is a review paper that summarizes the breeding techniques applied in tea such as conventional breeding, linkage mapping, association mapping and genomic selection. A proposed GS framework is also presented.

4**. Thematic area two:** This section has 3 papers, the current methods of tea quality evaluation; sensory, physical and chemical methods, and electronic techniques are described and compared. An optimized miniature process for use in tea breeding to manufacture black tea was evaluated. Lastly, biochemical analysis of 12 compounds was conducted using NMR and the results were correlated with sensory data to inform the key compounds for use to select high quality teas.

5**. Thematic area three:** Results on genomic prediction and genome wide association studies of quality traits in tea are presented. Significant markers and candidate genes associated with tea quality were identified. The accuracy of five GS models were compared using a 5-fold cross validation approach.

6. **General conclusion:** This section presents a general conclusion, based on the results of all the three thematic areas for all the objectives.

7. **Recommendation for future research:** This part suggests further research that could be performed to support and improve the results of this thesis.

List of papers

Table 1. List of draft manuscripts included in the main body of the thesis.

No.	Title	Status
1	Genomic selection and use of molecular approaches in tea (*Camellia sinensis* (L.) O. Kuntze) breeding: Present status and future prospects	Draft
2	Advances in techniques and methods for assessing tea (*Camellia sinensis* (L.) O. Kuntze) quality: A review of recent development	Draft
3	Evaluation of an optimized miniature process for use in black tea quality assessment in a tea breeding programme	Draft
4	Determinants of tea quality – multivariate analyses of green leaf biochemical compounds and organoleptic properties	Draft
5	Genomic prediction and genome wide association mapping for quality in tea (*Camellia sinensis* (L.) O. Kuntze) using genotyping-by-sequencing	Draft

General introduction

1.1. Tea breeding goals

The world's population is increasing steadily and is projected to reach 9 billion people in 2050 (Melorose et al., 2015), leading to an increase in the demand for food (Valin et al., 2014). Moreover, areas previously occupied by crops will be occupied by human beings. Tea is an important crop to the economies of many developing countries as it has led to creation of jobs for many people and development of good social infrastructure such as road networks, schools and hospitals in rural areas (Mukhtar et al., 2000b). For a sustained tea production, tea breeders are required to continuously release new superior varieties. The aim of crop improvement programmes is to make genetic gain over time for one or more traits of economic interest. Tea breeding goals vary amongst the major tea producing countries depending on the local needs (Mondal, 2014b; Jain et al., 2009). However, developing varieties with combined high yield and quality is the most important breeding objective (Mondal, 2014a; Kamunya et al., 2012). Currently, tea productivity is facing a serious threat of climate change, resulting in significant yield losses (Sitienei et al., 2017; Gunathilaka et al., 2017), and decreased quality (Han et al., 2017). Climate change has resulted in extreme, erratic and less predictable weather patterns, resulting in longer dry periods, heavy rainfall, more hail and increased temperatures (Marx et al., 2017; Batley et al., 2016). Additionally, the changing climate has led to increased attacks of pests and diseases (Ahmed et al., 2014). Therefore, most tea breeders are focusing on developing high yielding, high quality and varieties that are tolerant to biotic and abiotic stresses (Mondal, 2014b; 2011).

1.2. From conventional to molecular breeding of tea

Conventional tea breeding methods face several challenges including long gestation periods, high inbreeding depression, self-incompatibility, low success rate of hand pollination, short flowering time (2–3 months) of tea and the long duration for seed to mature (12–18 months) (Mondal, 2014b). It is therefore a time consuming and expensive process. Few studies have been conducted in tea using marker assisted

selection (MAS) techniques such as linkage mapping and association mapping (Koech et al., 2018b). MAS requires prior identification of the genes or markers which are significantly associated with the trait of interest (Semagn et al., 2010). Important traits in tea such as yield, quality and drought tolerance are quantitative in nature and are controlled by many genes, each with small effect that are difficult to identify using QTL mapping (Heffner et al., 2009b). With the current focus on improving breeding efficiency, reducing the cost of tea breeding and accelerating development of improved tea varieties, GS is a new approach that promises to improve gain from selection per unit time compared to phenotypic selection and the classical marker assisted selection methods (QTL mapping and association mapping) (De Donato et al., 2013; Heffner et al., 2010). GS uses genome-wide markers to predict the breeding values of individuals in a breeding population. The main advantage of GS is that it doesn't require the identification of QTLs or markers linked to the traits of interest, and hence its' suitable for selecting complex quantitative traits (Crossa et al., 2017; Goddard, 2009; Meuwissen et al., 2001). GS enables selection to be made at any time as long as the genotypic information is available and could potentially replace some field evaluation stages hence shortening the breeding cycles of tea.

References

Ahmed, S., Orians, C. M., Griffin, T. S., Buckley, S., Unachukwu, U., Stratton, A. E., Stepp, J. R., Robbat, A., Cash, S., Kennelly, E. J. (2014). Effects of water availability and pest pressures on tea (Camellia sinensis) growth and functional quality. *AoB Plants, 6*.

Batley, J., Edwards, D. (2016). The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Current Opinion in Plant Biology, 30*, 78-81.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science, 22*(11), 961-975.

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One, 8*(5), e62137.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica, 136*(2), 245-257.

Gunathilaka, R. P. D., Smart, J. C. R., Fleming, C. M. (2017). The impact of changing climate on perennial crops: the case of tea production in Sri Lanka. *Climatic change, 140*(3-4), 577-592.

Han, W.-Y., Huang, J.-G., Li, X., Li, Z.-X., Ahammed, G. J., Yan, P., Stepp, J. R. (2017). Altitudinal effects on the quality of green tea in east China: a climate change perspective. *European Food Research and Technology, 243*(2), 323-330.

Heffner, E. L., Sorrells, M. E., Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. *Crop Science, 49*(1). doi:10.2135/cropsci2008.08.0512
Heffner, E. L., Lorenz, A. J., Jannink, J.-L., Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Science, 50*(5), 1681-1690.
Jain, S. M., Priyadarshan, P. M. (2009). *Breeding plantation tree crops: tropical species*: Springer.
Kamunya, S. M., Wachira, F. N., Pathak, R. S., Muoki, R. C., Sharma, R. K. (2012). Tea improvement in Kenya. In *Global Tea Breeding* (pp. 177-226): Springer.
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 Genetics & Genomes, 14*(1). doi:10.1007/s11295-017-1219-8
Marx, W., Haunschild, R., Bornmann, L. (2017). Global warming and tea production—The bibliometric view on a newly emerging research topic. *Climate, 5*(3), 46.
Melorose, J., Perroy, R., Careas, S. (2015). World population prospects. *United Nations, 1*(6042), 587-592.
Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.
Mondal, T. K. (2014a). *Breeding and biotechnology of tea and its wild species*: Springer Science & Business Media.
Mondal, T. K. (2014b). Genetics and Breeding. In *Breeding and Biotechnology of Tea and its Wild Species* (pp. 9-34).
Mondal, T. K. (2011). Camellia. In *Wild crop relatives: genomic and breeding resources* (pp. 15-39): Springer.
Mukhtar, H., Ahmad, N. (2000). Tea polyphenols: prevention of cancer and optimizing health–. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.
Semagn, K., Bjørnstad, Å., Xu, Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology, 13*(5), 16-17.
Sitienei, B. J., Juma, S. G., Opere, E. (2017). On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County, Kenya. *Climate, 5*(3), 54.
Valin, H., Sands, R. D., Van der Mensbrugghe, D., Nelson, G. C., Ahammad, H., Blanc, E., Bodirsky, B., Fujimori, S., Hasegawa, T., Havlik, P. (2014). The future of food demand: understanding differences in global economic models. *Agricultural Economics, 45*(1), 51-67.

1.3. Aims and Objectives

The main aim of this research is to present a proof-of-concept for the applicability of GS and GWAS in tea breeding programmes. The specific objectives are structured around three broad thematic areas.

1.3.1. Thematic area one

Tea breeding in the modern era. Specific objectives under this thematic area are;

1. To review the conventional tea breeding process, its limitations and discuss current application of MAS in tea breeding.
2. To develop a GS framework for its implementation in tea breeding programmes.

1.3.2. Thematic area two

Tea quality evaluation methods. Specific objectives within this thematic area are;

1. To review the current tea quality evaluation techniques and make suggestions for future tea quality research in the tea industry.
2. To investigate the reliability of using an optimized miniature process for manufacturing newly developed varieties into black tea for assessment of their quality potential before release for commercial planting.
3. To evaluate the quality of tea genotypes using biochemical data obtained from NMR and determine the correlation between the measured green leaf biochemical compounds and organoleptic evaluation results.

1.3.3. Thematic area three

Genomic prediction and genome wide association studies of quality traits in tea. Specific objectives within this thematic area are;

1. Perform a GWAS to identify significant SNPs and identify potential candidate genes associated with tea quality.

2. To assess the effectiveness and potential of GBS as a genotyping platform for GS studies in tea breeding.
3. To estimate genomic and broad sense heritabilities of twelve quality traits in tea evaluated using Nuclear Magnetic Resonance spectroscopy (NMR).
4. To compare the prediction accuracies of five GS models: RRBLUP, BayesLASSO, BayesA, BayesB and BayesCπ.

Literature review

2.1. Botanical classification and the genome of tea

Tea (*Camellia sinensis* (L.) O. Kuntze) is a perennial crop (2n = 30) that belongs to the genus *Camellia* and family *Theaceae* (Namita et al., 2012). It is a diploid (2n=30) (Kondo, 1977), highly outcrossing, self-incompatible and highly heterozygous tree species (Muoki et al., 2007). Natural polyploids also exist in tea (Mondal, 2014a). There are three main varieties of cultivated tea worldwide. Camellia *sinensis* var. *sinensis* (CSS; Chinese type) and *Camellia sinensis* var. *assamica* (CSA; Assam type) are the two main types, both of which have recently been sequenced (Wei et al., 2018a; Xia et al., 2017a). *Camellia sinensis* var. *sinensis* has small semi-erect leaves and with a genome size of 3.1 Gb (Wei et al., 2018a), while *Camellia sinensis* var. *assamica* has large leaves and with a genome size of 3.02 Gb (Xia et al., 2017a). CSS grows slowly but is able to withstand colder climates, while CSA grows quickly and is highly sensitive to cold weather, and is mainly cultivated in warm tropical areas (Wei et al., 2018a). CSS is mainly grown in high altitude areas for the production of high quality green tea, while CSA is usually planted for use in processing black tea (Willson et al., 2012). The *assamica* tea originated from the forests of Assam in north-eastern India while the *sinensis* type is believed to originate from Sichuan province, south-western China (Van der Vossen et al., 2000). The third variety of tea is the *Camellia assamica* subsp. *Lasiocalyx*, that has medium-sized leaves and is commonly known as Cambod type (Ming, 2000).

2.2. Distribution and economic importance of tea

Currently, tea is grown in over 52 countries in tropical and subtropical regions around the world and is an important cash crop in many developing countries (Meegahakumbura et al., 2016). China and India are ranked first and second highest tea producing countries in the world and they account for 36.28% and 22.61% of the global tea production, respectively (FAO, 2018). Although Kenya is the third largest producer of tea in the world after China and India, it is also the world's leading exporter of black tea (ITC, 2018). China type teas are mainly cultivated across South China and in some Southeast Asian countries.

Assam is the most cultivated type and is grown widely in India and other tea growing countries across the globe (Meegahakumbura et al., 2016). The cambod tea was originally cultivated only in Indo-China (South Yunnan of China, Myanmar, Assam in India, Nothern Thailand, Vietnam, Laos and Cambodia), but currently is produced worldwide (Meegahakumbura et al., 2016).

Tea is a major source of livelihood in many countries, as it has led to the creation of jobs for many people and development of good social infrastructure such as road networks, schools and hospitals in rural areas (Mukhtar et al., 2000a). For instance, tea is the highest foreign exchange earner for Kenya. The tea sector contributed to about 4% of Kenya's GDP and 26% of the country's total export earnings in 2017 (TBK, 2018). In most developing countries, tea is mainly grown in rural areas, hence contributing to the improved living standard of the rural communities.

2.3. Types of tea

Processed tea is mainly classified according to the method of fermentation (Takeo, 1992) as black, green and oolong teas (Tao et al., 2016b). Black tea is made from leaves that are completely fermented by the deliberate aeration of the fresh green leaves. Fermentation is an enzymatic process whereby polyphenol oxidase catalyze the polyphenols present in green leaf into theaflavins and thearubigins (Friedman et al., 2005). Theaflavins and thearubigins are important compounds that give tea desirable quality attributes such as taste, mouthfeel, colour and aroma (Obanda et al., 2004). Black tea accounts for approximately 72% of the world's total tea production (Sharangi, 2009). Oolong tea is partially fermented while green tea is unfermented (Von Gadow et al., 1997). Other minor teas produced are white, yellow, Pu-erh and reprocessed tea (flower scented tea, compressed tea, instant tea and herbal teas) (Mishra et al., 2018a). The manufacturing process of white tea involves collecting the buds and young tea leaves shortly before the buds have fully opened. Then the leaves are steamed and dried with minimum amount of processing. For this reason, white tea retains the greatest levels of antioxidants and the lowest levels of caffeine compared to the major teas (green, black or oolong) (Sharangi, 2009). Pu-erh tea processing is like that of black or

green teas. However, the difference is, it is aged for some times to develop the flavor and colour of tea (Sharangi, 2009).

2.4. Tea quality and biochemical compounds influencing tea quality

Tea quality is described commercially by sensory characteristics such as colour, aroma, texture, taste of tea liquor and the appearance of dry tea (Zheng et al., 2016a). Flavour is comprised of taste, mouthfeel and aroma, while colour is the appearance of dry tea and liquor (Ho et al., 2015). Taste, mouthfeel and colour are formed by non-volatile compounds while aroma is produced by volatile compounds in tea (Liang et al., 2003). Studies conducted so far show that catechins, caffeine and amino acids present in fresh green tea influence the quality of black tea (Wright et al., 2000b). Several studies have demonstrated that tea with high catechins, caffeine and amino acids (theanine) content is of high quality (Owuor et al., 2007; Zuo et al., 2002; Wright et al., 2000b).

The main chemical compounds in fresh tea include polyphenols, amino acids, carbohydrates, alkaloids, minerals, vitamins, volatile compounds and pigments (Chlorophyll and carotenoids) (Caffin et al., 2004; Harbowy et al., 1997). Polyphenols consisting of flavonoids and phenolic acids are the main compounds known to influence tea quality (Owuor et al., 2007; Wright et al., 2000b; Obanda et al., 1997). Flavonoids are composed of flavan-3-ols (catechins), flavones, flavonols, flavanones, anthocyanins and isoflavonoids (Hodgson et al., 2010).

Flavan-3-ols (catechins) constitute between 60–80 % of the total polyphenols (Balentine et al., 1997) and they include (-)-epigallocatechin-3-gallate (EGCG), (-)-epigallocatechin (EGC), (-)-epicatechin-3- gallate (ECG) and epicatechin (EC) , (+)-gallocatechin (GC), (-)-gallocatechin gallate (GCG), (-)-catechin gallate (CG) and (+)-catechin (C) (El-Shahawi et al., 2012a). Catechins are generally water soluble and contribute to the bitterness, astringency and sweet aftertaste of tea brew (Liang et al., 2003). Catechins are broadly classified as non-gallated catechins (C, GC, EC and EGC) and gallated catechins (ECG and EGCG) (Peters et al., 2010). The gallated catechins are astringent and give tea the characteristic bitter taste while the non-gallated catechins (EGC and EC) are far less astringent (Xu et al., 2018b). Catechins are also

known to have therapeutic properties such as anticancer, antihypertension, anti-vascular and anti-inflammatory effects (Sharangi, 2009).

Flavonols consists of three main groups namely quercetin, kaempferol and myricetin together with their glycosides. They are present as free flavonols and as flavonol glycosides. The glycosidic group may be composed of glucose, galactose, rhamnose, rutin, and p-coumaric, and they contribute to bitterness, astringency and colour of tea (Jiang et al., 2015). Flavanones, flavones and isoflavones are present in low quantities and contribute to the taste and medicinal properties of tea (Wang et al., 2000b).

Anthocyanins consists of six main groups; cyanidin, delphinidin, malvidin, peodin, petunidin, and pelargonidin (Das et al., 2016). More than 500 different anthocyanins have been identified in plants based on the hydroxylation, methoxylation patterns on the B ring and glycosylation with different sugar units (Das et al., 2016). Anthocyaninin is mainly present in purple coloured tea, and the predominant type is malvidin (Kerio et al., 2012). Anthocyanins are important quality indicators in tea (Wrolstad et al., 2005). Similarly, anthocyanin have health benefits such as anti-inflammatory (Dai et al., 2007) antioxidant (Bae et al., 2007), antiartherosclerotic (Mazza, 2007), antimicrobial (Viskelis et al., 2009) and anti-carcinogenic effects (Wang et al., 2008c).

Phenolic acids are grouped into two classes which include hydroxybenzoic acids and hydroxycinnamic acids (Wang et al., 2000b). The main phenolic acids present in tea are gallic acid, p-coumaric acid, chlorogenic acid, benzoic acids and caffeic acid(van der Hooft et al., 2012). Phenolic acids contribute to the astringency properties of tea (Scharbert et al., 2004).

Alkaloids present in tea include caffeine, theophylline and of theobromine. Caffeine is the largest concentration (between 2% to 4% dry weight) while theophylline and theobromine are available in low concentrations at 0.1% (Ashihara et al., 1997). In addition to affecting the taste of tea with its sharp bitterness, caffeine has mood and cognitive enhancing properties and it is an essential component of tea which contributes to its quality (Caffin et al., 2004; Harbowy et al., 1997). The quality of tea positively correlates with the concentration of caffeine, by contributing to the astringency of black tea infusion. Moreover, caffeine removes fatigue and sleepy feeling and has diuretic action (Snel et al., 2011).

There are more than 600 volatile compounds identified and they contribute to the aroma of tea (Schuh et al., 2006). These compounds can broadly be classified into primary or secondary products (Ho et al., 2015). The primary products are biosynthesized by the tea plant and are present in the fresh green leaf and do not undergo changes. Secondary products are produced during tea manufacture from substrates such lipids, amino acids and carotenoids through enzymatic and non-enzymatic reactions (Ho et al., 2015). The primary volatile compounds in tea products are biosynthesized through six main pathways and they include carotenoid derivatives pathway, the fatty acid derivatives pathway, the terpene derivatives pathway, the phenylpropanoid/benzenoid derivatives pathway, the glycoside hydrolysis pathway and the Maillard reaction pathway (Zheng et al., 2016a; Ho et al., 2015).

The main pigments present in fresh tea are chlorophyll and carotenoids. Most of the chlorophyll is preserved during manufacture of green tea and hence the characteristic green colour (Suzuki et al., 2003). However, some chlorophyll is transformed to pheophytin and epimers due to the heat treatment during processing (Suzuki et al., 2003). In black tea, chlorophyll degrades and transforms to either pheophytin or chlorophyllide, catalyzed by chlorophylase and finally to pheophorbide. About 14 types of carotenoids have been identified in fresh shoots of tea (Suzuki et al., 2003). The main carotenoids in tea are β-carotene, lutein, zeaxanthin, neoxanthine, xanthophyll, lycopene and violaxanthine (Suzuki et al., 2003; Ravichandran, 2002). Carotenoids are major precursors for aroma formation, and therefore significantly contribute to tea quality (Zheng et al., 2016a; Ho et al., 2015). Both chlorophyll and carotenoids contribute to the colour and taste of made tea(Ravichandran, 2002).

2.5. Factors influencing tea quality

The genetic makeup of tea varieties has a major influence on tea quality. Different varieties exhibit varying levels of catechins and caffeine in tea, and can therefore be used as biochemical markers of tea quality (Owuor et al., 2007; Magoma et al., 2000). The quantities and proportion of these biochemical compounds in green leaf are positively correlated with black tea quality (Owuor et al., 2007). Flavour and colour of tea liquor and the appearance of dry leaf are influenced by polyphenols, catechins and caffeine

levels present in tea. Higher levels of these compounds in fresh shoots are preferable as they are indicators of high-quality tea. Therefore, the choice of the variety to cultivate is important as it determines the quality of tea (Dutta et al., 2011).

Environmental conditions have a significant impact on tea quality. Tea is mainly cultivated in tropical and subtropical regions because they have conducive environment for the growth of tea. The environmental factors influencing tea quality include soil type, altitude, seasons, weather factors (Owuor et al., 2011). Tea grown in high altitudes is of high quality because it is more aromatic compared to tea grown in low altitudes. Tea grows best at high and evenly distributed rainfall which range from 1150 to 1400 mm per year (Carr et al., 1992). In areas where rainfall is less than 1,150 mm per annum and those with long and hot dry spells, irrigation is recommended (Carr et al., 1992). Tea grows under a wide range of temperatures which range from 18 to 20 °C (Carr et al., 1992). Similarly, tea requires high relative humidity between 80–90%. Soil properties that lead to high tea yields and quality include a soil pH of between 4.0 and 5.5 (Goswami et al., 2001) and high organic matter (Ananthacumaraswamy et al., 2002).

Management practices such as fertilizer application, irrigation and harvesting operations significantly influence the quality of tea (Owuor et al., 2011). Harvesting tea by hand leads to high quality tea because the young tender shoots (2 leaves and a bud) are plucked at the right size. It is recommended to harvest tea at the correct interval. Long plucking interval leads to overgrown shoots which require course plucking, therefore introducing undesirable leaf quality. Course plucking leads to reduced levels of aroma, polyphenols and polyphenol oxidase (PPO) activity hence poor-quality teas (Owuor et al., 2011). Fertilizers are essential for growth and establishment of tea. However, fertilizers should be applied according to the recommended rates since high rates of nitrogen reduce black tea quality and does not increase yields (Owuor et al., 2011).

Processing parameters applied during tea manufacture have a significant impact on tea quality (Obanda et al., 2001). Black tea manufacture processes involve withering, cutting, curling, tear, fermentation and drying (Ravichandran, 2002). Optimization of each of the steps can affect the quality of tea (Ravichandran et al., 1998). Moisture content during withering, fermentation temperatures, relative

humidity and time of fermentation, influence the colour and liquor brightness (Owuor et al., 2017; Obanda et al., 2004).

2.6. Tea breeding

The rates of genetic gain delivered by plant breeding must be doubled if we are to achieve the productivity and sustainability targets of 2050 (Fischer et al., 2009). The main breeding approaches in tea include conventional breeding (controlled hybridization and individual selection) and molecular marker assisted selection (Chen et al., 2007b). The most important breeding objectives for tea include; high yield, high quality, high efficiency and high tolerance to biotic and abiotic stresses (Mondal, 2014a; Kamunya et al., 2012; Chen et al., 2007b).

2.6.1. Conventional tea breeding

Conventional breeding has been the main method used in selecting improved tea varieties in many countries worldwide (Mondal, 2014a; Kamunya et al., 2012). All the teas cultivated worldwide originated from India and China and were introduced to other countries (Meegahakumbura et al., 2016). Field selection and hybridization are the main approaches used in conventional tea breeding (Mondal, 2014a). Field selection involves identification of superior bushes in naturally existing seedling populations (Chen et al., 2007b). Promising bushes are selected and established as bi-clonal and polyclonal seed gardens (Carr, 2018). The resulting seedlings are used as parents, which have given rise to the current germplasm through crossing and selection (Meegahakumbura et al., 2018b; Kamunya et al., 2012). Hybridization and hand pollination are modern methods of tea variety development and they involve crossing selected tea types and wild relatives to produce new crop varieties based on Mendel's laws of inheritance (Carr, 2018; Sharma et al., 2010). Hybridization methods practiced in tea breeding are either natural or controlled (Mondal, 2014a; Chen et al., 2007b). Natural hybridization involves using bi-clonal or polyclonal seed baries that are planted in an isolated field, allowed to flower and cross pollinate naturally (Carr, 2018; Mondal, 2014a). Since tea is self-incompatible (Muoki et al., 2007; Wachira et al., 2005), any seed set is a cross between the clones

(Carr, 2018). F_1 seeds from these baries are harvested, planted and the superior seedlings are selected for important traits such as yield, quality and drought tolerance. The best seedlings are multiplied and exposed to multiple locations to test for genotype x environment effects (Kamunya et al., 2012). The most stable varieties across locations are multiplied and released as clones for commercial planting (Carr, 2018). Polyclonal seed baries involves planting several tea bushes together, and this introduces more variability among the F1 seeds. However, it is difficult to identify the parentage of offspring resulting from polyclonal seed baries, unless molecular markers are used (Hou et al., 2006; Tanaka et al., 2005a). In controlled hand pollination, two parents with known traits of interest are crossed manually by hand and subsequent generations examined for progenies with the desired traits (Carr, 2018). The main limitation of controlled hand pollination is that the success rate is very low (Carr, 2018), implying that many crosses are required, resulting in high costs and labour.

The standard method of developing superior tea varieties at Unilever Tea Kenya (UTK) is as described in Figure 1. Parents with desirable traits in tea (yield, quality, drought tolerance, pest and disease tolerance) are selected and crossed for one year. The seeds are germinated and allowed to grow in a tea nursery for one year, before being established in the field as hand pollinated trials (HPT). Hand pollinated trials are single progeny trials and each seedling is distinct from each other. Yield data from the HPTs are recorded for 3 years from each of the single bush progenies separately and the best 20% of the seedlings are selected, and vegetatively propagated in the tea nursery. The target is to get 24 uniform plants from each of the selected superior seedling. Mini clonal trial (MCT) is established as a completely randomized block design, consisting of 12 plants per clone replicated twice. Phenotypic evaluation for yield and other traits in the MCT is recorded for 5 years. Superior varieties are selected from the MCTs and vegetatively propagated in the nursery for one year. The last stage of selection in conventional tea breeding is clonal field trials (CFTs), and it involves clonal evaluation of superior varieties selected from the MCTs in a completely randomized block design, consisting of 16 plants per clone replicated twice. It is planted at different sites to test for the effects of genotype by environment interaction. CFTs are recorded for a further 5 years and the superior varieties are selected and bulked for 3 years, before being released for commercial

planting. The total duration of developing superior tea varieties at Unilever Tea Kenya is 16 years (Figure 1).

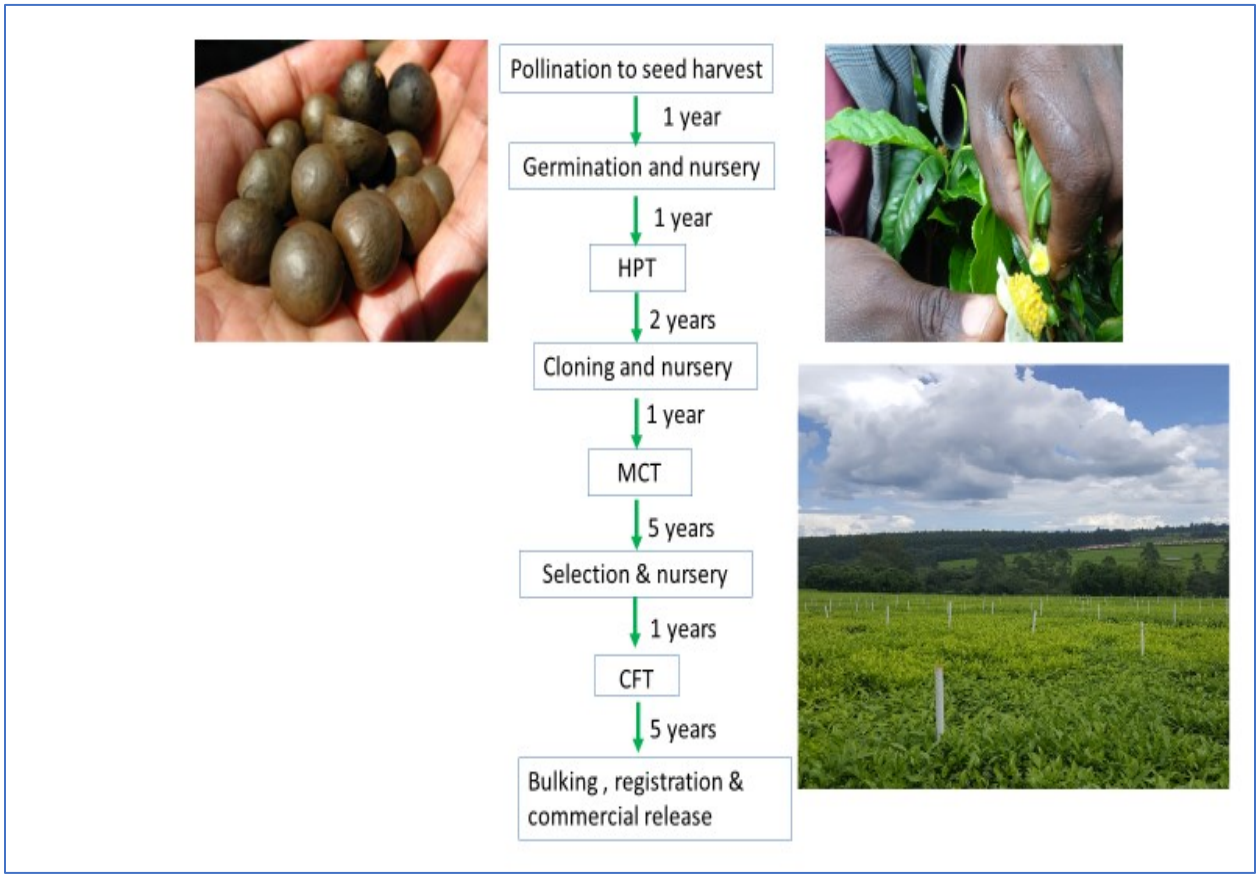

Figure 1. Conventional tea breeding at Unilever Tea Kenya Limited (HPT-hand pollinated trial, MCT-mini clonal trial, CFT- Clonal field trial).

2.6.2. Application of genetic markers in tea breeding

Genetic markers are genes or DNA sequences with known chromosome locations controlling a specific gene or trait and reveal genetic differences between individuals (Xu, 2010a; Semagn et al., 2006a). Genetic markers used in tea breeding are broadly grouped into classical markers and DNA/molecular markers (Mukhopadhyay et al., 2016). Classical markers are further divided into morphological, cytological and biochemical markers (Mukhopadhyay et al., 2016).

2.6.2.1. Morphological markers used in tea

Morphological markers are visually detectable plant characteristics such as the shape, size, colour and surface of various plant parts and have been used by tea breeders worldwide to select superior varieties (Mondal, 2014a). Tea taxonomists use morphological traits such as colour and shape of the leaf, growth habit and features of flowers to identify different tea species (Banerjee, 1992). Similarly, tea breeders frequently use characteristics such as bush vigour, recovery from prune, table height, dry matter production, shoot density and shoot replacement ratio to select high yielding tea varieties (Carr, 2018; Mukhopadhyay et al., 2016). Additionally, leaf pubescence (Kamunya et al., 2010) and pigmentation (Banerjee, 1992) are used as indicators for selecting high quality black tea varieties. Morphological markers are simple and inexpensive to use, since they do not require specialized biochemical and molecular techniques. However, they have not been used extensively in tea breeding because they are limited in number, require many mapping populations and are influenced by the environment and hence may lead to inaccurate identification of phenotypes (Kamunya et al., 2010).

2.6.2.2. Biochemical markers

Biochemical markers are variations in amino acid and protein banding pattern and are separated by molecular weight or isoelectric point on electrophoresis gels. Catechin content are used as biochemical markers in diversity studies of tea to select and differentiate tea genotypes (Magoma et al., 2000). Similarly, several classes of biochemical compounds such as volatiles, amino acids, total polyphenols, soluble sugars and caffeine in fresh green leaf are the main determinants of quality in black tea (Owuor et al., 2007; Wright et al., 2000b). Additionally, Nyarukowa et al. (2016) developed a simple method of screening drought tolerant tea varieties using amino acid and sugars. Biochemical markers are codominant and can distinguish both homozygotes and heterozygotes. However, biochemical markers are limited in number and are affected by the environment (Collard et al., 2005), and therefore molecular markers are preferred in tea breeding (Mukhopadhyay et al., 2016).

2.6.2.3. Cytological markers

Cytological markers are variations present in the numbers, banding patterns, size, shape, order and position of chromosomes. Cytological markers have been studied extensively in tea and karyotypic data for various species of tea is available (Mondal, 2011; Chen et al., 2000). Kondo et al. (1981) reported karyotypic variability among several species of tea using the C-banding method. However, determination of karyotype grouping in tea is difficult because of the high stickiness of the chromosomes. Similarly, the homologous pairs of chromosomes in tea are not identical (Kondo, 1977).

2.6.3. DNA markers/ molecular markers

DNA markers are specific DNA sequences on a chromosome associated with specific traits and reveal polymorphisms between individuals. Molecular markers are divided into hybridization-based markers, PCR-based markers and sequence or chip-based markers (Jiang, 2013; Semagn et al., 2006a). DNA Markers should ideally be; (1) highly polymorphic, to distinguish many individuals, (2) highly abundant and uniformly distributed in the genome, (3) co-dominant, to distinguish heterozygotes and homozygotes individuals, (4) reproducible, (5) cost-efficient for screening a large population and (6) easy to genotype (Semagn et al., 2006a; Gupta et al., 1999).

2.6.3.1. Restriction Fragment Length Polymorphism (RFLP)

RFLP is the only molecular marker based on hybridization and it relies on the ability of restriction endonucleases to recognize and cleave DNA at specific sites within the genome, generating many DNA fragments of different lengths (Winter et al., 1995). RFLP markers are codominant in nature, do not require sequence information and have relatively high reproducibility (Semagn et al., 2006a). RFLP markers have been utilized in genetic diversity studies to determine genetic relationship of Korean and Japanese teas (Matsumoto et al., 2004). These studies showed that the genetic diversity among Japanese tea is narrow, and the Korean tea germplasm was largely introduced from China and partially from Japan. Devarumath et

al. (2002) used RFLP markers to investigate the genetic integrity of micro-propagated diploid and triploid elite teas. They found that there were genetic variations at the DNA sequence level and clonal teas derived from organized meristems are not always genetically true to the type. Matsumoto et al. (2004) used RFLP markers to successfully distinguish Assam hybrids and Japanese green tea cultivars with high and low catechins content, and to classify different Japanese cultivars. Similarly, RFLP markers have been used to differentiate teas originated from China and Japan (Matsumoto et al., 2002). There are few practical applications of RFLP markers in tea compared to other markers because they are expensive, labour intensive, require large quantity of DNA and the steps involved in genotyping are tedious, complex and time consuming (Kamunya et al., 2010).

2.6.3.2. PCR based markers

2.6.3.2.1. Random amplified polymorphic DNA

RAPD markers detects DNA polymorphism produced by rearrangement or deletions at or between oligonucleotide primer binding sites in the genome (Williams et al., 1990). RAPD markers in tea have mainly been utilized in genetic diversity and population differentiation (Mukhopadhyay et al., 2016). For example, RAPD markers were used to estimate genetic diversity and taxonomic relationships in 38 clones of three different species of tea namely, *Camellia assamica*, *Camellia sinensis*, and *Camellia assamica* ssp. *Lasiocalyx* (Wachira et al., 1995). Large genetic variations were reported within and between the three species. Similar findings have also been reported in several other studies utilizing RAPD markers (Wachira et al., 2001; Kaundun et al., 2000). These results were confirmed by other findings that found tea to be highly outcrossing (Muoki et al., 2007; Wachira et al., 2005) and heterozygous crop (Banerjee, 1992). In other findings, Roy et al. (2009) used RAPD markers to determine genetic diversity among the China, Assam and Cambod species. They found that the China variety had the highest genetic diversity. Other studies using RAPD markers agreed with these findings (Chen et al., 2002). Likewise, RAPD markers have been used to determine the genetic fidelity of micro-propagated tea derived from ex-plants of field grown

mother bushes as well as in vitro germinated seedlings (Borchetia et al., 2009). These results revealed that tea grown using axillary and adventitious methods of propagation are genetically true to type. Other application of RAPD markers in tea includes parental identification (Wu et al., 2002), discrimination of genotypes (Chen et al., 2002), molecular phylogeny studies (Chen et al., 2000), assessment of mating system (Muoki et al., 2007), construction of linkage maps (Hackett et al., 2000). Additionally, Mphangwe et al. (2013), identified associations between RAPD markers and 6 desirable traits in tea traits namely black tea quality, high yield and tolerance to drought, high temperature, low temperature and Phomopsis theae. RAPD markers have widely been used in many crops including tea because they are technically simple and fast to assay, require small quantities of DNA, are highly polymorphic, are easily detected on ethidium bromide stained agarose gels and no prior knowledge about the genome is required (Semagn et al., 2006a). However, RAPD are dominant markers, not reliable as they are not easily reproducible within and between laboratories, hence not suitable for MAS and genetic mapping (Avise, 2012).

2.6.3.2.2. Amplified Fragment Length Polymorphism (AFLP)

AFLP are dominant and multi-locus fingerprinting markers that combines the reliability of the RFLP technique with the power of the PCR technique, hence they are more robust and reliable (Bensch et al., 2005; Liu et al., 2004). Several genetic studies have been conducted using AFLP markers in tea. Paul et al. (1997) used AFLP markers to detect diversity and genetic differentiation among Indian and Kenyan populations of tea (*Camellia sinensis* (L.) O. Kuntze). They reported that Kenyan tea germplasm is predominantly made up of genotypes originating from India, and the China varieties had the greatest genetic variation. Similar findings were revealed by Mishra et al. (2004) who showed that genetic diversity in tea varied according to origin and China varieties had the greatest diversity followed by Assam and Cambod types, respectively. Similar findings were also reported by Balasaravanan et al. (2003) who used AFLP markers to characterize 49 tea cultivars in India and they showed that varieties from China had the highest genetic diversity. In other studies, Sharma et al. (2010) found a high level of genetic variation (85%) within populations and low variation (15%) between common commercial varieties in India using AFLP markers.

Studies conducted by Raina et al. (2012) in India using AFLP markers reported that there were wide variations among tea accessions while commercial varieties had limited variabilities. Additionally, several genetic linkage maps for important traits in tea have been constructed using AFLP markers (Chang et al., 2017; Huang et al., 2005). AFLP markers are reliable and reproducible (Blears et al., 1998), easy to use compared to RFLPs (Valsangiacomo et al., 1995), do not require prior knowledge of the DNA sequence and only a small quantity of genomic DNA is required (Semagn et al., 2006a). However, AFLPs are dominant markers, require high molecular weight DNA and are expensive and labour intensive to develop (Semagn et al., 2006a).

2.6.3.2.3. Simple Sequence Repeat (SSR)

SSR are short tandem repeats of DNA sequence of about 2-6 bases in length and are widespread in all prokaryotic and eukaryotic chromosomes (Chistiakov et al., 2006; Zane et al., 2002). SSR markers are advantageous because they are co-dominant, ubiquitous, multi-allelic, highly reproducible and requires low amounts of DNA. However, the development of SSRs is tedious, costly and requires extensive knowledge of DNA sequence information (Semagn et al., 2006a).

SSR markers have been used widely in tea breeding. Studies conducted in Western Himalaya, India using SSR markers revealed moderate to high genetic diversity for superior clones commercially grown and could therefore be used in genetic improvement of tea (Bhardwaj et al., 2014). Ujihara et al. (2009) successfully used SSR markers to distinguish 16 Japanese genotypes and one Chinese tea cultivar. They concluded that SSR markers could be used for variety identification. Similar studies using SSR markers have been conducted to identify the origin of tea germplasm in India and China (Meegahakumbura et al., 2016). Wambulwa et al. (2016) used SSR markers to determine the genetic diversity and relationships of cultivars of African tea. They reported that the genetic diversity present in Africa tea is low and *Camellia sinensis* var. *assamica*, which originated from India is the main variety cultivated. Other studies conducted in East Africa using SSR markers also revealed *Camellia sinensis* var. *assamica* as the most popular variety, and it had the lowest genetic diversity compared to other varieties (Meegahakumbura et al., 2018b). A core

collection of tea was selected from worldwide germplasm resources using SSR markers (Taniguchi et al., 2014). These studies revealed that Japanese tea had the lowest genetic diversity while genotypes from China, Taiwan, India and Sri Lanka had the highest genetic diversity in that order (Taniguchi et al., 2014). SSR markers have also been used to construct genetic linkage maps for important traits in tea (Tan et al., 2016a; Lorenz et al., 2011b).

2.6.3.3. Sequence/ chip-based markers

2.6.3.3.1. Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are variations at a given position in a DNA sequence among individuals. In tea, SNP markers have been used to determine genetic diversity between *Camellia sinensis* and its wild relatives (Yang et al., 2016). They reported that cultivated accessions had greater heterozygosity than wild accessions, except *C. taliensis var. bangwei*, which was identified as semi-wild in the same study. The first reference map using SNP markers was constructed by Ma et al. (2015). A total of 6,448 SNPs were used and the map had fifteen linkage groups, with a total map length of 3,965 cM, and an average inter-locus distance of 1.0 cM. This map provides valuable information that can be utilized by tea breeders and geneticists in fine mapping of quantitative trait loci (QTL), map-based cloning and marker-assisted selection, hence improving desirable traits in tea (Ma et al., 2015). SNP markers have been used significantly in genetic studies, because of their high abundance, stability, bi-allelic variation in the genome of diploid species, and ease of automation in high-throughput analysis (Mochida et al., 2004). Although most common in non-coding regions, SNPs also occur in functional genomic regions and are therefore important for characterizing genes associated with complex traits (McCouch et al., 2010; Nelson et al., 2004).

2.6.3.3.2. Diversity arrays technology markers (DArT)

The diversity Array Technology (DArT) is a high throughput micro-array hybridization-based marker system that allows simultaneous detection of thousands of DNA polymorphisms arising from single base changes, and small insertions and deletions in a single assay (Wenzl et al., 2004). Compared to other marker systems, DArT markers are highly reproducible and provide a high throughput hence allowing for rapid analysis of many DNA samples in a single assay (Akbari et al., 2006). Additionally, DArT markers are cost effective (Xia et al., 2005) and do not need prior sequence information (Jaccoud et al., 2001). DArT markers have not been used widely in tea and only one application of DArT markers has so far been reported. Koech et al. (2018a) used DArT marker for mapping black tea quality and drought tolerance. The map consisted of 15 linkage groups and with a total length of 1260.1 cM and a mean interval of 1.1 cM between markers (Koech et al., 2018a).

2.7. Marker assisted selection (MAS)

Marker assisted selection involves the identification of genetic markers linked to specific genes of interest and utilization of these markers for indirect selection of desirable phenotypic traits. Before marker assisted selection is utilized in plant breeding, molecular markers influencing important traits are identified through genetic mapping and validated (Collard et al., 2007). Genetic maps are essential tools for implementing quantitative trait loci (QTL) analysis and marker-assisted selection (MAS) (Xu et al., 2017). Genetic mapping techniques include linkage mapping and association mapping (Xu et al., 2017). The aim of conducting genetic mapping is to understand the inheritance and genetic architecture of important quantitative traits and identify desirable markers for use in marker assisted selection (Collard et al., 2007). The markers identified in genetic mapping studies are validated before use in marker assisted selection (Collard et al., 2005).

Compared to conventional breeding, marker assisted selection is more reliable since it's not influenced by the environment. MAS is also cost-effective especially when evaluating traits such as yield, quality and resistance to biotic and abiotic stress which require expensive screening techniques (Collard et

al., 2007). Similarly, MAS reduces the time required to release varieties since selection using markers can be carried out early in the breeding cycle before the trait is expressed. Additionally, MAS is more efficient because markers linked to the trait of interest are selected and used, allowing the selection of traits that are not easily manipulated in conventional breeding (Collard et al., 2005). MAS also avoid the transfer of undesirable or deleterious genes especially during gene introgression from wild species to elite varieties (Ben-Ari et al., 2012). MAS is also effective in selecting traits with low heritability (Collard et al., 2005).

2.7.1. QTL mapping

QTL mapping is the determination of associations between specific DNA markers and phenotypic traits of interest in a segregating bi-parental population (Collard et al., 2005). The main steps involved in QTL mapping include; 1) developing an appropriate mapping population, 2) phenotyping the population for the trait of interest, 3) identification of the type of molecular markers for genotyping, 4) linkage analysis to identify markers linked to the trait of interest using statistical programs, and 5) marker validation to test the applicability and reliability of the identified markers associated with the QTLs in predicting the trait of interest (Semagn et al., 2010). The validation process of the markers identified involves testing the markers' effectiveness in different genetic backgrounds or independent breeding populations (Zhou et al., 2003; Cakir et al., 2003).

The construction of genetic linkage maps is the basis of molecular biology and is essential for a wide range of genetics and genomic studies, such as quantitative trait mapping, molecular marker-assisted breeding and comparative genomic studies (Xia et al., 2020). The first linkage map in tea was constructed by Hackett et al. (2000), using AFLP and RAPD markers. The map had a length of 1349.7 cM and 126 markers with an average distance of 11.7 cM between loci and 15 linkage groups, corresponding to the haploid number of tea. However, the map was assembled using markers from female parents only and therefore majority of the markers exhibited unexpected segregation ratios. Later, Huang et al. (2005) constructed an AFLP linkage map for tea using both female and male parents. The map of a female parent had 17 linkage groups and 208 markers, covering length of 2457.7 cM, and the average distance between

markers was 11.9 cM. AFLP linkage map for the male parent had 16 linkage groups and 200 markers, covering a total length of 2545.3 cM, and the average distance between markers was 12.8 cM (Huang et al., 2005). Kamunya et al. (2010) developed a linkage map using RAPD, AFLP and SSR markers tea to identify markers associated with yield in tea. The map contained 30 (19 maternal and 11 paternal) linkage groups with a total length of 1,411.5 cM and a mean interval of 14.1 cM between loci. Twenty-three putative yield QTLs were detected in the 2 environments studied. However, these studies used a small mapping population size comprising of only 42 progenies and therefore further validation steps were recommended before utilizing the markers in marker assisted breeding. The first high-density reference linkage map of tea was constructed using SSR markers (Taniguchi et al., 2012). The core map had 15 linkage groups, with a total length of 1218 cM. The combined maps had 441 SSRs, 7 CAPS, 2 STS and 674 RAPDs and was recommended for use as a basic reference linkage map of tea (Taniguchi et al., 2012). A moderately saturated genetic map was constructed using 406 SSR markers using a pseudo-testcross population of 183 individuals derived from an intraspecific cross of two *Camellia sinensis* varieties with diverse catechins composition (Ma et al., 2014). The map consisted of fifteen linkage groups with a total map length of 1,143.5 cM, and an average locus spacing of 2.9 cM. A total of 25 QTLs associated with catechins content were identified (Ma et al., 2014). Similarly, Tan et al. (2016a) constructed a saturated linkage map using an F1 tea population derived from crossing two varieties namely 'Longjin43' × 'Baihaozao'. The map consists of 15 linkage groups, covering a total length of 1226.2 cM with an average marker distance of 2.5 cM. A total of 15 QTLs were identified and were associated with timing of spring bud flush, young shoot colour, mature leaf length, mature leaf width and leaf shape index (Tan et al., 2016a). Recently a moderately saturated genetic map with a length of 1441 cM was constructed using RAPD, AFLP and SSR markers, and it provides a foundation for developing markers linked to disease resistance and tea quality (Chang et al., 2017). In this study, Japanese and Korean cultivars were used to make double pseudo-testcrosses of 79 F1 mapping population.

2.7.1.1. Limitations of QTL mapping

Developing a mapping population is time-consuming, tedious and an expensive process (Heffner et al., 2009b). It involves crossing two genetically divergent parents, differing in the traits of interest (Collard et al., 2005). However, the success rate of crossing tea is always low since many flowers abort (Carr, 2018), hence it is recommended to ensure that many crosses are done. Similarly, the self-incompatible nature of tea plants, leads to a low seed yield, leading to the propagation of insufficient populations for QTL mapping (Xia et al., 2020). Tea is also a perennial crop with a long generation time; it takes between 3 to 6 years for tea to grow from seed to flowering (Mondal, 2014a). This makes linkage mapping in tea is a time-consuming process.

The available resources for conducting QTL mapping limit the size of the mapping population to be used, which significantly affects the accuracy of detecting QTL positions and effect estimates (Heffner et al., 2009b; Schön et al., 2004). Land is a scarce, expensive and treasured resource in many countries, and this limits the number of linkage mapping experiments that can be conducted in tea. In addition, it is expensive to employ skilled labour specifically for hand pollination and maintaining seedlings in the nursery and in the field. Moreover, a marker validation step is required in QTL mapping, which is expensive and could limit the effectiveness of biparental population for MAS in plant breeding (Holland, 2004).

Biparental populations used in QTL mapping are small and only capture a small portion of the total genetic variation, hence do not represent the whole breeding population (Xu, 2010a). Studies conducted in tea, have shown that the small population used in QTL mapping leads to overestimated and spurious QTL effects (Kamunya et al., 2010). Additionally, the statistical methods used in linkage analysis are not adequate for improving quantitative traits, that are controlled by many genes each with small effect (Lorenz et al., 2011b).

2.7.2. Association mapping

Association mapping is a population-based method used to detect and map QTLs based on significant association of molecular markers and phenotypic traits (Gupta et al., 2005). Compared to linkage

analysis which utilizes data within a family to predict correlations between a phenotype and a marker, association mapping exploits the historical LD to identify trait-marker relationships within a natural non-biparental population (Pasam et al., 2012). Linkage disequilibrium is used to construct LD maps, study marker-trait association (MTA) both independently and in combination with linkage analysis and in population genetics and evolution studies (Gupta et al., 2005). Association mapping results in higher mapping resolution and evaluates a wide range of alleles rapidly compared to linkage mapping (Yu et al., 2006b; Stich et al., 2005).

The two main types of association mapping are candidate gene association mapping and genome wide association studies (GWAS). Candidate gene approach involves determination of associations between polymorphic variants in known genes of interest and quantitative traits. It requires knowledge of the location and function of genes that influence the genetic, biochemical and physiological pathways of the trait of interest (Hall et al., 2010; Mackay, 2001). Genome wide association studies involves scanning markers across the entire genome for significant statistical associations between a set of molecular markers and the phenotypic trait of interest. It requires the use of high marker density to accurately identify markers that reveal genome-wide LD structure and haplotype diversity in the study population (Zhu et al., 2008). The main steps in an association mapping studies include; 1) Selection of individuals from a natural population or germplasm collection with a wide genetic diversity, 2) phenotyping, 3) genotyping, 4) estimation of linkage disequilibrium, 4) assessment of population structure and kinship, 5) association analysis, and 6) replication and validation of the predicted QTLs (Abdurakhmonov et al., 2008).

Association mapping was first applied in human genetics to study associations between QTLs and diseases (Hindorff, 2009). Currently, association mapping methods are being applied in many crops (Soto-Cerda et al., 2012a; Hall et al., 2010; Sorrells et al., 2009). However, very few association mapping studies in tea have been conducted so far. Jin et al. (2016) conducted an association analysis and identified 4318 SNPs associated with caffeine content in four different environments. Significant marker–trait association was also validated, and the SNPs identified were recommended for use in marker-assisted selection for tea quality improvement (Jin et al., 2016). Similarly, Su et al. (2016) studied association between quality traits

in tea and markers in Southern Henan, China. They detected 2211 pairs of loci of which 259 had significant linkage disequilibrium with the traits (D′>0.5). 19 EST-SSR markers were found to be significantly associated with polyphenols, total free amino acids and caffeine. The findings from this work can be used for improvement of tea quality and early identification of varieties in tea breeding programmes before undergoing field trials which take a long time (Su et al., 2016). With the availability of next generation sequencing technologies at much reduced costs, more marker- trait association studies in tea can now be conducted.

2.7.2.1. Advantages of association mapping over QTL mapping

Association mapping is conducted on the existing natural population that is readily available, and therefore identification of QTLs does not involve developing a bi-parental population which is expensive and time-consuming. Similarly, there are higher levels of genetic recombination within an association mapping population, which leads to a higher resolution compared to the bi-parental population applied in QTL mapping (Korte et al., 2013a). Additionally, many alleles can be evaluated simultaneously for one locus, and this ensures effective gene tagging and accurate identification of markers for use in marker assisted selection. Association mapping is also more cost effective compared to QTL mapping since many traits can be investigated using the same panel of accessions and genotypic data without the need for an expensive validation procedure (Breseghello et al., 2006).

2.7.2.2. Limitations of association studies

Although association leads to a greater resolution compared to QTL mapping, it often leads to overestimation of effects due to the arbitrary significance thresholds set, which sometimes leads to poor trait predictions (Korte et al., 2013a). Similarly, the population structure and kinship within an association mapping population leads to spurious associations, hence predicting false positive marker-trait association (Zhao et al., 2007a). In addition, the detection of marker-trait association using association mapping is

influenced by allele frequency distribution within a population and therefore, it's difficult to identify a causative rare allele (Myles et al., 2009), leading to biased estimation of effect and poor prediction of performance of individuals (Rafalski, 2002).

2.8. Genomic selection (GS)

Genomic selection is a type of marker assisted selection that simultaneously estimates all locus, haplotype and marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs) for use in trait prediction (Hickey et al., 2014; Heffner et al., 2009b; Meuwissen, 2001). It uses high density markers covering the whole genome so that all quantitative trait loci (QTLs) affecting traits of interest are in linkage disequilibrium with at least one molecular marker (Goddard et al., 2007). Unlike linkage mapping and association mapping that considers only significant QTLs, GS incorporates all markers in a population regardless of their effect at each locus, by summing all the marker effects to predict breeding values of individuals (Lorenz et al., 2011b; Heffner et al., 2009b). This makes GS a promising approach for predicting complex traits such as yield, quality and drought tolerance which are influenced by many genes with small effects (Jannink et al., 2010a).

Genomic selection utilizes a training and breeding population. The training population consists of individuals genotyped using high density genome wide markers and phenotyped for traits of interest. The breeding or candidate population consists of individuals with only genotypic data. The genotypic and phenotypic data obtained from the training population is used to build genomic selection models for predicting genomic estimated breeding values of the selection candidates. The model captures total additive genetic variance across the entire genome to estimate the genomic estimated breeding values (GEBV) of individuals in the breeding population based on the sum of all marker effects (Hickey et al., 2014). GEBVs are the predicted value of selection candidates based on marker effects (Goddard et al., 2007). Genome wide markers are treated as random effects and all marker effects on the phenotype are estimated simultaneously in a single model. The superior individuals from the candidate population are selected based on the predicted values (GEBVs) (Hayes et al., 2001).

Before genomic selection models are used in breeding programmes, they are validated to determine the prediction accuracy. Prediction accuracy is the correlation between GEBVs and the true breeding values, estimated by measuring phenotypic performance (Zhong et al., 2009a; Goddard et al., 2007). Genomic selection in crops is projected to reduce cost of breeding (Jannink et al., 2010a), enhance breeding efficiency by increasing prediction accuracy, increase selection intensity and benefit from rapid gain from selection (Muranty et al., 2015; Grattapaglia, 2014). There are currently no reports of the application of GS in tea breeding.

2.8.1. Statistical Methods in genomic selection

Genomic selection models utilize linear mixed models to estimate GEBVs while the traditional marker assisted selection methods exploits regular linear models. Traditional MAS considers markers as fixed effects and requires a stepwise regression approach to fit the markers singly or in small groups. Marker effects are set at zero or to their full value depending on whether their significance is below or above the predetermined threshold. Therefore, the marker effects are usually overestimated in QTL mapping and association mapping (Xu, 2003).

Genomic selection was developed to overcome the shortcomings of linkage mapping and association mapping by minimizing the biased marker effects estimation (Goddard et al., 2007). The markers are treated as predictor variables (p) and the phenotypic measurements as response variables (n). Molecular markers are usually more than the phenotypic measurements, implying that there are more predictor variables compared to phenotypic observations, hence creating a “large p small n problem” (Lorenz et al., 2011b). The degrees of freedom are not adequate for estimating all predictor effects simultaneously using least squares approach. Even if there were sufficient degrees of freedom, a high degree of multicollinearity among markers may occur leading to an overfitted model that exaggerates minor fluctuations, and this consequently reduces the prediction accuracy (Lorenz et al., 2011b; Jannink et al., 2010a). Therefore, a variety of statistical models have been developed to solve the problem of having large markers and few phenotypes, by estimating the effect of each marker (SNP effects) to predict important

traits. These models differ in the assumptions on variance of marker effect and the type of gene action (Lorenz et al., 2011b). The main statistical models used include ridge regression best unbiased linear predictor (RR-BLUP), which is equivalent to the genomic best unbiased linear predictor (G-BLUP) and Bayesian models (BayesA, BayesB, BayesC, Bayes LASSO) (Lorenz et al., 2011b). The basic GS model is expressed using the following equation (Moser et al., 2009);

$$Yi = g(xi) + ei$$

Where yi is an observed phenotype of individual i (i = 1 . . . n) and xi is a 1 x p vector of SNP genotypes on individual i, g(xi) is a function relating genotypes and phenotypes, and ei is a residual term. GEBV is generally equal to g(xi).

2.8.1.1. Ridge regression best linear unbiased prediction (RR-BLUP)

This method is also known as random regression best linear unbiased prediction (RR-BLUP). It assumes that marker effects are random and have a normal distribution with common variance resulting in equal shrinkage of their effects towards zero (Hayes et al., 2001). This means that the amount of shrinkage is the same for all markers (Heffner et al., 2009b; Bernardo et al., 2007). However, the assumption that individual markers have the same variance is not realistic for many traits exploited in plant breeding and may lead to over-shrinking of large effects genes (Xu, 2003). Despite this limitation, RR-BLUP has been used extensively in plant breeding and the results have been accurate and reliable (Grattapaglia et al., 2018; Lipka et al., 2014).

RR-BLUP model is mathematically equivalent to genomic best linear unbiased prediction model (GBLUP), which has low computational requirements and is easy to use (Endelman, 2011; VanRaden et al., 2009). GBLUP is a linear mixed model which incorporates a marker-based genomic relationship matrix (GRM), generated by evaluating marker covariance across all individuals (Endelman, 2011; Hayes et al., 2009a). GS models that utilize GRM have better resolutions of genetic relationships among individuals and result in higher accuracy compared to pedigree selection methods in plants (Crossa et al., 2010). Moreover,

models that apply GRM can be applied to individuals with no breeding history or pedigree records (Hayes et al., 2013).

2.8.1.2. Bayesian Estimation models

In Bayesian models, a separate variance is estimated for each marker, and the variances are assumed to follow a specified prior distribution, resulting in unequal shrinkage of their effects (Lorenz et al., 2011b; Hayes et al., 2001). Bayesian models include Bases A, Bayes B, Bayes Cπ and Bayes LASSO. Bayes A is also known as Bayesian shrinkage regression model (Xu, 2003) and was originally proposed by Meuwissen (2001). In this method, markers are assumed to have different variances and are modelled following a scaled inverse $\chi 2$ distribution (Meuwissen, 2001). It assigns t-distribution for marker effects which causes strong shrinkage towards zero for small estimates of marker effects and less shrinkage for sizable estimates of marker effects.

Bayes B was developed by Meuwissen (2001). It assumes that marker effects follow a t-distribution, and most loci have no effect on the trait and hence most markers are left out of the prediction model. The Bayes B models presumes that the trait is controlled by few loci that vary in effect size (Lorenz et al., 2011b). These markers are represented by the point mass at var (β) = 0, and therefore allows markers with no effects to be incorporated. Similarly, markers included in the model have effects sampled from distributions with different variances. This model is more realistic because some regions of the genome are not associated with QTLs affecting a trait and therefore have a zero-effect estimate (Meuwissen, 2001). Compared to other Bayesian methods, Bayes B is more accurate and is also less computationally demanding (Heffner et al., 2009b).

Bayes Cπ method is similar to Bayes B except that it assigns normal distribution to nonzero marker effects. Marker effects included in the model are sampled from the same distribution whose variance is estimated from the data. This gives Bayes Cπ more flexibility to model oligogenic to polygenic traits (Lorenz et al., 2011b).

Bayesian least absolute shrinkage and selection operator (Bayes LASSO) has an exponential prior on the marker variances giving a double exponential distribution for the marker effects resulting in more shrinkage for small marker effects and less shrinkage for large effects. It allows some predictor variables to have variance equal to zero and performs continuous shrinkage of the remaining variables simultaneously (Lorenz et al., 2011b). This makes Bayes LASSO a suitable model when major genes are present. However, it only selects n predictor variables at most and does not perform well when the variables are correlated.

2.8.2. Factors affecting the accuracy of genomic selection models

In genomic selection, a model that produces high accuracy (r) is desirable. Statistically, accuracy is the degree of similarity between the true value and an estimated value (VanRaden et al., 2009). In genomic selection, accuracy of selection is the correlation between genomic estimated breeding values (GEBVs) and the true breeding value (Lorenz et al., 2011b). The true breeding value is the estimated breeding value (EBV) and is predicted using traditional methods that utilize phenotypic data (Asoro et al., 2011b; Heffner et al., 2009b). This correlation provides an estimate of selection accuracy and directly relates GEBV prediction accuracy to response in selection (Asoro et al., 2011b). GS models are evaluated by measuring the prediction accuracy derived from the breeder's equation; $R = ir\sigma A/T$, where R is the response from selection, i is the selection intensity, r is the selection accuracy, σA is the square root of the additive genetic variance and T is the length of time to complete a breeding cycle (Falconer et al., 1996).

In Gs, one or more markers should be in linkage disequilibrium with every gene affecting the trait of interest (Hickey et al., 2014), and therefore an increase in the number of LD steadily improves the prediction accuracy (Asoro et al., 2011b). Similarly, the marker type and density influence the prediction accuracy (Hayes et al., 2013). Studies have shown that dense genome wide markers are preferable to cover as many QTLs as possible, and therefore increasing the prediction accuracy (Crossa et al., 2010; Meuwissen et al., 2009). The level of LD between a marker and a QTL can be used to determine the optimum marker density required to achieve a high prediction accuracy (Lorenz et al., 2011b). This implies that more markers are required if the linkage disequilibrium between SNPs and QTL is not adequate.

A large training population is required to accurately estimate marker effects, which lead to improved prediction accuracies (Jannink et al., 2010a). Meuwissen (2001) reported a positive significant correlation between the training population size and prediction accuracy. Other studies have shown that increasing training population size increased the prediction accuracy by 20% (Lorenzana et al., 2009). Similarly, the training population used in training the GS models should be closely related to the breeding population (Crossa et al., 2017; Hickey et al., 2014).

Trait heritability is a key factor that impacts on the accuracy of GS (Heffner et al., 2011). Traits with high heritability increases the GS accuracy (Riedelsheimer et al., 2012). The heritability of a trait significantly affects the response to selection and improves the efficiency of GS over phenotypic selection (Zhang et al., 2017a; Hayes et al., 2009a). High heritability leads to increased gain from selection for the traits of interest (Lorenz et al., 2011b).

References

Abdurakhmonov IY, Abdukarimov A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International journal of plant genomics, 2008*.

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, et al. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and applied genetics, 113*(8), 1409-1420.

Ananthacumaraswamy S, Singh ID. (2002). An overview on the factors affecting the quality of Sri Lankan black tea. *Journal of food science and technology, 39*(5), 449-457.

Ashihara H, Gillies FM, Crozier A. (1997). Metabolism of caffeine and related purine alkaloids in leaves of tea (Camellia sinensis L.). *Plant and cell physiology, 38*(4), 413-419.

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The plant genome, 4*(2), 132-144.

Avise JC. (2012). *Molecular markers, natural history and evolution*: Springer Science & Business Media.

Bae S-H, Suh H-J. (2007). Antioxidant activities of five different mulberry cultivars in Korea. *LWT-Food Science and Technology, 40*(6), 955-962.

Balasaravanan T, Pius PK, Kumar RR, Muraleedharan N, Shasany AK. (2003). Genetic diversity among south Indian tea germplasm (Camellia sinensis, C. assamica and C. assamica spp. lasiocalyx) using AFLP markers. *Plant Science, 165*(2), 365-372.

Balentine DA, Wiseman SA, Bouwens LCM. (1997). The chemistry of tea flavonoids. *Critical Reviews in Food Science & Nutrition, 37*(8), 693-704.

Banerjee B. (1992). Selection and breeding of tea. In *Tea* (pp. 53-86): Springer.

Ben-Ari G, Lavi U. (2012). Marker-assisted selection in plant breeding. In *Plant Biotechnology and Agriculture* (pp. 163-184): Elsevier.

Bensch S, Åkesson M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology, 14*(10), 2899-2914.

Bernardo R, Yu J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop science, 47*(3), 1082-1090.
Bhardwaj P, Sharma RK, Kumar R, Sharma H, Ahuja PS. (2014, 2014). *SSR marker based DNA fingerprinting and diversity assessment in superior tea germplasm cultivated in Western Himalaya*.
Blears MJ, De Grandis SA, Lee H, Trevors JT. (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology, 21*(3), 99-114.
Borchetia S, Das SC, Handique PJ, Das S. (2009). High multiplication frequency and genetic stability for commercialization of the three varieties of micropropagated tea plants (Camellia spp.). *Scientia Horticulturae, 120*(4), 544-550.
Breseghello F, Sorrells ME. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop science, 46*(3), 1323-1330.
Caffin N, D'Arcy B, Yao L, Rintoul G. (2004). Developing an index of quality for Australian tea, Rural Industries Research and Development Corporation. In.
Cakir M, Gupta S, Platz GJ, Ablett GA, Loughman R, Emebiri LC, et al. (2003). Mapping and validation of the genes for resistance to Pyrenophora teres f. teres in barley (Hordeum vulgare L.). *Australian Journal of Agricultural Research, 54*(12), 1369-1377.
Carr MKV. (2018). *Advances in tea agronomy*: Cambridge University Press.
Carr MKV, Stephens W. (1992). Climate, weather and the yield of tea. In *Tea* (pp. 87-135): Springer.
Chang Y, Oh EU, Lee MS, Kim HB, Moon D-G, Song KJ. (2017). Construction of a genetic linkage map based on RAPD, AFLP, and SSR markers for tea plant (Camellia sinensis). *Euphytica, 213*(8), 190.
Chen L, Wang P-S, Satoshi Y. (2002). Discrimination of wild tea germplasm resources (Camellia sp.) using RAPD markers. *Agricultural Sciences in China, 1*(10), 1105-1110.
Chen L, Yu F, Tong Q. (2000). Discussions on phylogenetic classification and evolution of Sect. Thea. *Journal of Tea Science, 20*(2), 89-94.
Chen L, Zhou Z-X, Yang Y-J. (2007). Genetic improvement and breeding of tea plant (Camellia sinensis) in China: from individual selection to hybridization and molecular breeding. *Euphytica, 154*(1-2), 239-248.
Chistiakov DA, Hellemans B, Volckaert FAM. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture, 255*(1-4), 1-29.
Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196.
Collard BCY, Mackill DJ. (2007). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1491), 557-572.
Crossa J, de Los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics, 186*(2), 713-724.
Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in plant science, 22*(11), 961-975.
Dai J, Patel JD, Mumper RJ. (2007). Characterization of blackberry extract and its antiproliferative and anti-inflammatory properties. *Journal of medicinal food, 10*(2), 258-265.
Das B, Patel PK, Sobhapandit S, Gogoi RC. (2016). Screening and determination of anthocyanin in pigmented tea germplasm. *Two and a Bud, 63*(1), 1-3.
Devarumath R, Nandy S, Rani V, Marimuthu S, Muraleedharan N, Raina S. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing Camellia sinensis (China type) and C. assamica ssp. assamica (Assam-India type). *Plant cell reports, 21*(2), 166-173.

Dutta R, Stein A, Bhagat RM. (2011). Integrating satellite images and spectroscopy to measuring green and black tea quality. *Food chemistry, 127*(2), 866-874.
El-Shahawi MS, Hamza A, Bahaffi SO, Al-Sibaai AA, Abduljabbar TN. (2012). Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography. *Food chemistry, 134*(4), 2268-2275.
Endelman JB. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The plant genome, 4*(3), 250-255.
Falconer DS, Mackay TFC. (1996). Introduction to quantitative genetics. Longman. *Essex, England*.
FAO. (2018). FAOSTAT: Download data. Statistics Division, Food and Agriculture Organization of the United Nations,Rome, Italy.
Fischer RA, Byerlee DR, Edmeades GO. (2009). *Can technology deliver on the yield challenge to 2050?* Retrieved from
Friedman M, Kim SY, Lee SJ, Han GP, Han JS, Lee KR, et al. (2005). Distribution of catechins, theaflavins, caffeine, and theobromine in 77 teas consumed in the United States. *Journal of food science, 70*(9), C550-C559.
Goddard ME, Hayes BJ. (2007). Genomic selection. *Journal of Animal breeding and Genetics, 124*(6), 323-330.
Goswami H, Barooah AK, Gohain KK, Saikia BP, Baruah A, Dutta A. (2001). Soil parameters for higher productivity of tea. *Two and a Bud, 48*(1), 26-30.
Grattapaglia D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. In *Genomics of plant genetic resources* (pp. 651-682): Springer.
Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, et al. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in plant science, 9*.
Gupta PK, Rustgi S, Kulwal PL. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant molecular biology, 57*(4), 461-485.
Gupta PK, Varshney RK, Sharma PC, Ramesh B. (1999). Molecular markers and their applications in wheat breeding. *Plant breeding, 118*(5), 369-390.
Hackett CA, Wachira FN, Paul S, Powell W, Waugh R. (2000). Construction of a genetic linkage map for Camellia sinensis (tea). *Heredity, 85*(4), 346.
Hall D, Tegström C, Ingvarsson PK. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in Functional Genomics, 9*(2), 157-165.
Harbowy ME, Balentine DA, Davies AP, Cai Y. (1997). Tea chemistry. *Critical reviews in plant sciences, 16*(5), 415-480.
Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution, 41*(1), 51.
Hayes BJ, Cogan NOI, Pembleton LW, Goddard ME, Wang J, Spangenberg GC, et al. (2013). Prospects for genomic selection in forage plant species. *Plant breeding, 132*(2), 133-143.
Hayes BJ, Goddard ME. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.
Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop science, 51*(6), 2597-2606.
Heffner EL, Sorrells ME, Jannink J-L. (2009). Genomic selection for crop improvement. *Crop science, 49*(1), 1-12.
Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, et al. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop science, 54*(4), 1476-1488.
Hindorff LA. (2009). A catalog of published genome-wide association studies. *http://www. genome. gov/26525384*.
Ho C-T, Zheng X, Li S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9-27.
Hodgson JM, Croft KD. (2010). Tea flavonoids and cardiovascular health. *Molecular aspects of medicine, 31*(6), 495-502.

Holland JB. (2004, 2004). *Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities*.

Hou YJ, He Q, Liang GL, Li PW, Peng P, Deng M. (2006). ISSR analysis of the hybrid of the descendants of tea camellias. *Journal of Southwest Agricultural University (Natural Science), 28*(2), 267-270.

Huang JA, Li JX, Huang YH, Luo JW, Gong ZH, Liu ZH. (2005). Construction of AFLP molecular markers linkage map in tea plant. *J Tea Sci, 25*(1), 7-15.

ITC. (2018). International Tea Committee-Annual Bulletin of Statistics. . *International Tea Committee.*

.

Jaccoud D, Peng K, Feinstein D, Kilian A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic acids research, 29*(4), e25-e25.

Jannink J-L, Lorenz AJ, Iwata H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics, 9*(2), 166-177.

Jiang G-L. (2013). Molecular markers and marker-assisted breeding in plants. *Plant breeding from laboratories to fields*, 45-83.

Jiang H, Engelhardt UH, Thräne C, Maiwald B, Stark J. (2015). Determination of flavonol glycosides in green tea, oolong tea and black tea by UHPLC compared to HPLC. *Food chemistry, 183*, 30-35.

Jin J-Q, Yao M-Z, Ma C-L, Ma J-Q, Chen L. (2016). Association mapping of caffeine content with TCS1 in tea plant and its related species. *Plant physiology and biochemistry, 105*, 251-259.

Kamunya SM, Wachira FN, Pathak RS, Korir R, Sharma V, Kumar R, et al. (2010). Genomic mapping and testing for quantitative trait loci in tea (Camellia sinensis (L.) O. Kuntze). *Tree genetics & genomes, 6*(6), 915-929.

Kamunya SM, Wachira FN, Pathak RS, Muoki RC, Sharma RK. (2012). Tea improvement in Kenya. In *Global Tea Breeding* (pp. 177-226): Springer.

Kaundun SS, Zhyvoloup A, Park Y-G. (2000). Evaluation of the genetic diversity among elite tea (Camellia sinensis var. sinensis) accessions using RAPD markers. *Euphytica, 115*(1), 7-16.

Kerio LC, Wachira FN, Wanyoko JK, Rotich MK. (2012). Characterization of anthocyanins in Kenyan teas: Extraction and identification. *Food chemistry, 131*(1), 31-38. doi:https://doi.org/10.1016/j.foodchem.2011.08.005

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Apostolides Z. (2018). Identification of novel QTL for black tea quality traits and drought tolerance in tea plants (Camellia sinensis). *Tree genetics & genomes, 14*(1), 9.

Kondo K. (1977). Chromosome numbers in the genus Camellia. *Biotropica*, 86-94.

Kondo K, Parks CR. (1981). Cytological studies in cultivated species of Camellia. VI. Giemsa C-banded karyotypes of seven accessions of Camellia japonica L. sensu lato. *Japanese Journal of Breeding, 31*(1), 26-34.

Korte A, Farlow A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods, 9*(1), 29.

Liang Y, Lu J, Zhang L, Wu S, Wu Y. (2003). Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions. *Food chemistry, 80*(2), 283-290.

Lipka AE, Lu F, Cherney JH, Buckler ES, Casler MD, Costich DE. (2014). Accelerating the switchgrass (Panicum virgatum L.) breeding cycle using genomic selection approaches. *PloS one, 9*(11), e112227.

Liu ZJ, Cordes JF. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1-4), 1-37.

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. (2011). Genomic selection in plant breeding: knowledge and prospects. In *Advances in agronomy* (Vol. 110, pp. 77-123): Elsevier.

Lorenzana RE, Bernardo R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and applied genetics, 120*(1), 151-161.

Ma J-Q, Huang L, Ma C-L, Jin J-Q, Li C-F, Wang R-K, et al. (2015). Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PloS one, 10*(6), e0128798.

Ma J-Q, Yao M-Z, Ma C-L, Wang X-C, Jin J-Q, Wang X-M, et al. (2014). Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (Camellia sinensis). *PloS one, 9*(3), e93131.

Mackay TFC. (2001). The genetic architecture of quantitative traits. *Annual review of genetics, 35*(1), 303-339.

Magoma GN, Wachira FN, Obanda M, Imbuga M, Agong SG. (2000). The use of catechins as biochemical markers in diversity studies of tea (Camellia sinensis). *Genetic Resources and Crop Evolution, 47*(2), 107-114.

Matsumoto S, Kiriiwa Y, Takeda Y. (2002). Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA. *Theoretical and applied genetics, 104*(6-7), 998-1002.

Matsumoto S, Kiriiwa Y, Yamaguchi S. (2004). The Korean tea plant (Camellia sinensis): RFLP analysis of genetic diversity and relationship to Japanese tea. *Breeding science, 54*(3), 231-237.

Mazza G. (2007). Anthocyanins and heart health. *Annali-Istituto Superiore Di Sanita, 43*(4), 369.

McCouch SR, Zhao K, Wright M, Tung C-W, Ebana K, Thomson M, et al. (2010). Development of genome-wide SNP assays for rice. *Breeding science, 60*(5), 524-535.

Meegahakumbura MK, Wambulwa MC, Li M-M, Thapa KK, Sun Y-S, Möller M, et al. (2018). Domestication origin and breeding history of the tea plant (Camellia sinensis) in China and India based on nuclear microsatellites and cpDNA sequence data. *Frontiers in plant science, 8*, 2270.

Meegahakumbura MK, Wambulwa MC, Thapa KK, Li MM, Möller M, Xu JC, et al. (2016). Indications for three independent domestication events for the tea plant (Camellia sinensis (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PloS one, 11*(5), e0155369.

Meuwissen THE, Hayes, B. J., Goddard, M. E. . (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.

Meuwissen THE, Solberg TR, Shepherd R, Woolliams JA. (2009). A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genetics Selection Evolution, 41*(1), 2.

Ming TL. (2000). Monograph of the Genus Camellia. Yunnan Science and Technology Press. *Kunming, China*, 128-134.

Mishra P, Nordon A, Tschannerl J, Lian G, Redfern S, Marshall S. (2018). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering, 238*, 70-77. doi:https://doi.org/10.1016/j.jfoodeng.2018.06.015

Mishra RK, Sen-Mandi S. (2004). Molecular profiling and development of DNA marker associated with drought tolerance in tea clones growing in Darjeeling. *Current Science*, 60-66.

Mochida K, Yamazaki Y, Ogihara Y. (2004). Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetics and Genomics, 270*(5), 371-377.

Mondal TK. (2014). *Breeding and biotechnology of tea and its wild species*: Springer Science & Business Media.

Mondal TK. (2011). Camellia. In *Wild crop relatives: genomic and breeding resources* (pp. 15-39): Springer.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genetics Selection Evolution, 41*(1), 56.

Mphangwe NIK, Vorster J, Steyn JM, Nyirenda HE, Taylor NJ, Apostolides Z. (2013). Screening of tea (Camellia sinensis) for trait-associated molecular markers. *Applied biochemistry and biotechnology, 171*(2), 437-449.

Mukhopadhyay M, Mondal TK, Chand PK. (2016). Biotechnological advances in tea (Camellia sinensis [L.] O. Kuntze): a review. *Plant cell reports, 35*(2), 255-287.

Mukhtar H, Ahmad N. (2000). Tea polyphenols: prevention of cancer and optimizing health. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.

Muoki RC, Wachira FN, Pathak RS, Kamunya SM. (2007). Assessment of the mating system of Camellia sinensis in biclonal seed orchards based on PCR markers. *The Journal of Horticultural Science and Biotechnology, 82*(5), 733-738.
Muranty H, Troggio M, Sadok IB, Al Rifaï M, Auwerkerken A, Banchi E, et al. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture research, 2*, 15060.
Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, et al. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell, 21*(8), 2194-2202.
Namita P, Mukesh R, Vijay KJ. (2012). Camellia sinensis (green tea): A review. *Global journal of pharmacology, 6*(2), 52-59.
Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, et al. (2004). Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome research, 14*(8), 1664-1668.
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. *Journal of plant physiology, 198*, 39-48.
Obanda M, Owuor PO, Mang'oka R. (2001). Changes in the chemical and sensory quality parameters of black tea due to variations of fermentation time and temperature. *Food chemistry, 75*(4), 395-404.
Obanda M, Owuor PO, Mang'oka R, Kavoi MM. (2004). Changes in thearubigin fractions and theaflavin levels due to variations in processing conditions and their influence on black tea liquor brightness and total colour. *Food chemistry, 85*(2), 163-173.
Obanda M, Owuor PO, Taylor SJ. (1997). Flavanol composition and caffeine content of green leaf as quality potential indicators of Kenyan black teas. *Journal of the Science of Food and Agriculture, 74*(2), 209-215.
Owuor PO, Kamau DM, Kamunya SM, Msomba SW, Uwimana MA, Okal AW, et al. (2011). Effects of genotype, environment and management on yields and quality of black tea. In *Genetics, Biofuels and Local Farming Systems* (pp. 277-307): Springer.
Owuor PO, Obanda M. (2017). Clonal Variations in the Response of Hard Physically Withered Leaf to Rehydration Following Long Chemical Wither Durations.
Owuor PO, Obanda M. (2007). The use of green tea (Camellia sinensis) leaf flavan-3-ol composition in predicting plain black tea quality potential. *Food chemistry, 100*(3), 873-884.
Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC plant biology, 12*(1), 16.
Paul S, Wachira FN, Powell W, Waugh R. (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (Camellia sinensis (L.) O. Kuntze) revealed by AFLP markers. *Theoretical and applied genetics, 94*(2), 255-263.
Peters CM, Green RJ, Janle EM, Ferruzzi MG. (2010). Formulation with ascorbic acid and sucrose modulates catechin bioavailability from green tea. *Food Research International, 43*(1), 95-102.
Rafalski JA. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science, 162*(3), 329-333.
Raina SN, Ahuja PS, Sharma RK, Das SC, Bhardwaj P, Negi R, et al. (2012). Genetic structure and diversity of India hybrid tea. *Genetic Resources and Crop Evolution, 59*(7), 1527-1541.
Ravichandran R. (2002). Carotenoid composition, distribution and degradation to flavour volatiles during black tea manufacture and the effect of carotenoid supplementation on tea quality and aroma. *Food chemistry, 78*(1), 23-28.
Ravichandran R, Parthiban R. (1998). The impact of processing techniques on tea volatiles. *Food chemistry, 62*(3), 347-353.
Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics, 44*(2), 217.
Roy SC, Chakraborty BN. (2009). Genetic diversity and relationships among tea (Camellia sinensis) cultivars as revealed by RAPD and ISSR based fingerprinting.

Scharbert S, Holzmann N, Hofmann T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *Journal of agricultural and food chemistry, 52*(11), 3498-3508.
Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics, 167*(1), 485-498.
Schuh C, Schieberle P. (2006). Characterization of the key aroma compounds in the beverage prepared from Darjeeling black tea: quantitative differences between tea leaves and infusion. *Journal of agricultural and food chemistry, 54*(3), 916-924.
Semagn K, Bjørnstad Å, Ndjiondjop MN. (2006). An overview of molecular marker methods for plants. *African journal of biotechnology, 5*(25).
Semagn K, Bjørnstad Å, Xu Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology, 13*(5), 16-17.
Sharangi AB. (2009). Medicinal and therapeutic potentialities of tea (Camellia sinensis L.)–A review. *Food Research International, 42*(5-6), 529-535.
Sharma RK, Negi MS, Sharma S, Bhardwaj P, Kumar R, Bhattachrya E, et al. (2010). AFLP-based genetic diversity assessment of commercially important tea germplasm in India. *Biochemical genetics, 48*(7-8), 549-564.
Snel J, Lorist MM. (2011). Effects of caffeine on sleep and cognition. In *Progress in brain research* (Vol. 190, pp. 105-117): Elsevier.
Sorrells ME, Yu J. (2009). Linkage disequilibrium and association mapping in the Triticeae. In *Genetics and genomics of the Triticeae* (pp. 655-683): Springer.
Soto-Cerda BJ, Cloutier S. (2012). Association mapping in plant genomes. *Genetic Diversity in Plants*, 29-54.
Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. (2005). Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theoretical and applied genetics, 111*(4), 723-730.
Su H, Liu J, He W, Wen B, Li M, Feng J. (2016). Association analysis of traits related to tea (Camellia sinensis) quality with EST-SSRs in Southern Henan area. *Journal of Agricultural Biotechnology, 24*(9), 1328-1336.
Suzuki Y, Shioi Y. (2003). Identification of chlorophylls and carotenoids in major teas by high-performance liquid chromatography with photodiode array detection. *Journal of agricultural and food chemistry, 51*(18), 5307-5314.
Takeo T. (1992). Green and semi-fermented teas. In *Tea* (pp. 413-457): Springer.
Tan L-Q, Wang L-Y, Xu L-Y, Wu L-Y, Peng M, Zhang C-C, et al. (2016). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (Camellia sinensis). *Tree genetics & genomes, 12*(3), 52.
Tanaka T, Mizutani T, Shibata M, Tanikawa N, Parks CR. (2005). Cytogenetic studies on the origin of Camellia× vernalis. V. Estimation of the seed parent of C.× vernalis that evolved about 400 years ago by cpDNA analysis. *Journal of the Japanese Society for Horticultural Science, 74*(6), 464-468.
Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, et al. (2012). Construction of a high-density reference linkage map of tea (Camellia sinensis). *Breeding science, 62*(3), 263-273.
Taniguchi F, Kimura K, Saba T, Ogino A, Yamaguchi S, Tanaka J. (2014). Worldwide core collections of tea (Camellia sinensis) based on SSR markers. *Tree genetics & genomes, 10*(6), 1555-1565.
Tao W, Zhou Z, Zhao B, Wei T. (2016). Simultaneous determination of eight catechins and four theaflavins in green, black and oolong tea using new HPLC–MS–MS method. *Journal of pharmaceutical and biomedical analysis, 131*, 140-145.
TBK. (2018). Tea industry performance in 2018. *The Tea Board of Kenya.*
Ujihara T, Ohta R, Hayashi N, Kohata K, Tanaka J-i. (2009). Identification of Japanese and Chinese green tea cultivars by using simple sequence repeat markers to encourage proper labeling. *Bioscience, biotechnology, and biochemistry, 73*(1), 15-20.

Valin H, Sands RD, Van der Mensbrugghe D, Nelson GC, Ahammad H, Blanc E, et al. (2014). The future of food demand: understanding differences in global economic models. *Agricultural Economics, 45*(1), 51-67.

Valsangiacomo C, Baggi F, Gaia V, Balmelli T, Peduzzi R, Piffaretti J-C. (1995). Use of amplified fragment length polymorphism in molecular typing of Legionella pneumophila and application to epidemiological studies. *Journal of Clinical Microbiology, 33*(7), 1716-1719.

van der Hooft JJJ, Akermi M, Ünlü FY, Mihaleva V, Roldan VG, Bino RJ, et al. (2012). Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts. *Journal of agricultural and food chemistry, 60*(36), 8841-8850.

Van der Vossen HAM, Umali BE. (2000). *Plant resources of South-East Asia*: Backhuys Publ.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science, 92*(1), 16-24.

Viskelis P, Rubinskienė M, Jasutienė I, Šarkinas A, Daubaras R, Česonienė L. (2009). Anthocyanins, antioxidative, and antimicrobial properties of American cranberry (Vaccinium macrocarpon Ait.) and their press cakes. *Journal of food science, 74*(2), C157-C161.

Von Gadow A, Joubert E, Hansmann CF. (1997). Comparison of the antioxidant activity of rooibos tea (Aspalathus linearis) with green, oolong and black tea. *Food chemistry, 60*(1), 73-77.

Wachira F, Tanaka J, Takeda Y. (2001). Genetic variation and differentiation in tea (Camellia sinensis) germplasm revealed by RAPD and AFLP variation. *The Journal of Horticultural Science and Biotechnology, 76*(5), 557-563.

Wachira FN, Kamunya SK. (2005). Pseudo-self-incompatibility in some tea clones (Camellia sinensis (L.) O. Kuntze). *The Journal of Horticultural Science and Biotechnology, 80*(6), 716-720.

Wachira FN, Waugh R, Powell W, Hackett CA. (1995). Detection of genetic diversity in tea (Camellia sinensis) using RAPD markers. *Genome, 38*(2), 201-210.

Wambulwa MC, Meegahakumbura MK, Chalo R, Kamunya S, Muchugi A, Xu JC, et al. (2016). Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree genetics & genomes, 12*(1), 11.

Wang H, Provan GJ, Helliwell K. (2000). Tea flavonoids: their functions, utilisation and analysis. *Trends in Food Science & Technology, 11*(4), 152-160. doi:https://doi.org/10.1016/S0924-2244(00)00061-3

Wang L-S, Stoner GD. (2008). Anthocyanins and their role in cancer prevention. *Cancer letters, 269*(2), 281-290.

Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, et al. (2018). Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proceedings of the National Academy of Sciences, 115*(18), E4151-E4158.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, et al. (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences, 101*(26), 9915-9920.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research, 18*(22), 6531-6535.

Willson KC, Clifford MN. (2012). *Tea: Cultivation to consumption*: Springer Science & Business Media.

Winter P, Kahl G. (1995). Molecular marker technologies for plant improvement. *World Journal of Microbiology and Biotechnology, 11*(4), 438-448.

Wright LP, Mphangwe NIK, Nyirenda HE, Apostolides Z. (2000). Analysis of caffeine and flavan-3-ol composition in the fresh leaf of Camellia sinesis for predicting the quality of the black tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 80*(13), 1823-1830.

Wrolstad RE, Durst RW, Lee J. (2005). Tracking color and pigment changes in anthocyanin products. *Trends in Food Science & Technology, 16*(9), 423-428.

Wu LJ, Peng SZ, Xian Li J, Wen SC, Huan HY, Hua GZ, et al. (2002). Studies on genetic diversity of tea cultivars [Camellia sinensis (L.) O. Kuntze] by RAPD analysis. *J Tea Sci, 12*, 121-127.

Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Molecular plant, 10*(6), 866-877.

Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M, et al. (2005). DArT for high-throughput genotyping of cassava (Manihot esculenta) and its wild relatives. *Theoretical and applied genetics, 110*(6), 1092-1098.

Xu S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics, 163*(2), 789-801.

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 chemistry, 258*, 16-24.

Xu Y. (2010). *Molecular plant breeding*: Cabi.

Xu Y, Li P, Yang Z, Xu C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal, 5*(2), 175-184.

Yang H, Wei C-L, Liu H-W, Wu J-L, Li Z-G, Zhang L, et al. (2016). Genetic divergence between Camellia sinensis and its wild relatives revealed via genome-wide SNPs from RAD sequencing. *PloS one, 11*(3), e0151424.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics, 38*(2), 203.

Zane L, Bargelloni L, Patarnello T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology, 11*(1), 1-16.

Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in plant science, 8*, 1916.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS genetics, 3*(1), e4.

Zheng X-Q, Li Q-S, Xiang L-P, Liang Y-R. (2016). Recent advances in volatiles of teas. *Molecules, 21*(3), 338.

Zhong S, Dekkers JCM, Fernando RL, Jannink J-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics, 182*(1), 355-364.

Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK, Smith NJ. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant breeding, 122*(1), 40-46.

Zhu C, Gore M, Buckler ES, Yu J. (2008). Status and prospects of association mapping in plants. *The plant genome, 1*(1), 5-20.

Zuo Y, Chen H, Deng Y. (2002). Simultaneous determination of catechins, caffeine and gallic acids in green, Oolong, black and pu-erh teas using HPLC with a photodiode array detector. *Talanta, 57*(2), 307-316.

Genomic selection and use of molecular approaches in tea (*Camellia sinensis* (L.) O. Kuntze) breeding: Present status and future prospects

Lubanga M. Nelson[b,c*], De-Silva Jacquie[d], Massawe Festo[b], Mayes Sean[a]

*nelsonlubanga@gmail.com

[a]School of Biosciences, The University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK,

[b]University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[c]Unilever Tea Kenya Ltd., Nakuru Road, Kericho, 20-20200, Kenya

[d]Unilever Research and Development, Colworth Science Park, Sharnbrook, Bedford MK441LQ, UK

Abstract

Conventional tea breeding has been successful at delivering improved varieties using field selection and hybridization for many years. However, the long tea breeding process, emerging pests and diseases and climate change pose formidable challenges. A few studies aimed at accelerating tea breeding and improving selection precision through marker assisted selection (MAS) have been conducted in the past. However, linkage mapping and association mapping have not been successful in effectively driving practical MAS in tea breeding, because most of the important traits such as yield, quality, and tolerance to drought are quantitative in nature with a complex inheritance pattern. Genomic selection (GS) can address some limitations of conventional tea breeding and classical marker assisted selection techniques. GS uses whole genome-wide prediction models to estimate genome estimated breeding values (GEBVs) of a candidate population. With the decreased cost of genotyping and advanced genotyping tools, GS adoption in tea breeding has the potential to shorten the length of developing superior varieties, increase selection intensity and improve the accuracy of selection. The large amount of phenotypic and genomic information in many tea breeding programmes and development of many computational methods provide opportunities to

enhance the adoption of GS in tea breeding. GS has been applied successfully in animal breeding and in several crops such as wheat, maize, oil palm and forest trees with promising results. However, it has not been applied in tea improvement. In this paper, we review the conventional tea breeding process and its limitations and discuss current application of marker assisted selection in tea breeding. We then propose a GS framework/strategy for its implementation in tea breeding programmes.

Key words; Conventional tea breeding, marker assisted selection, genomic selection

3.1. Introduction

Tea (*Camellia sinensis* (L) O. Kuntze) is a perennial tree plant (2n = 30) that belongs to the genus *Camellia* in the family *Theaceae* (Chang, 1981). It is an important crop in many developing countries located mainly in Asia, Africa and South America, due to its' economic, medicinal and cultural values (Mondal, 2014b; Chen et al., 2012). The two main varieties of tea cultivated worldwide are *Camellia sinensis* var. *sinensis* (CSS; Chinese type) and *Camellia sinensis* var. *assamica* (CSA; Assam type) (Yao et al., 2012a), both of the genomes have recently been sequenced (Wei et al., 2018b; Xia et al., 2017b). *Camellia sinensis* var. *sinensis* has a genome size of 3.1 Gb (Wei et al., 2018b), while *Camellia sinensis* var. *assamica* has a 3.02 Gb genome size (Xia et al., 2017b). Tea is a diploid (2n=30) (Kondo, 1977), self-incompatible and hence highly heterozygous plant (Muoki et al., 2007; Wachira et al., 2005). Natural polyploids in tea also exist (Devarumath et al., 2002; Wachira et al., 1991).

Conventional tea breeding has been practised since 1939 (Mondal, 2014a) and has resulted in the development of many improved tea varieties (Mukhopadhyay et al., 2016). However, it is labour intensive, expensive, time consuming, influenced by the environment and has low selection accuracy (Xu et al., 2018a; Mukhopadhyay et al., 2016; Mondal, 2014b). Plant breeding of major crops has moved from being completely based on phenotypic selection to genotype-based selection due to the improvement in molecular biology and high-throughput genotyping technologies (Leng et al., 2017; Varshney et al., 2014).

Linkage mapping (Ye et al., 2018; Koech et al., 2018b) as well as association mapping (Jin et al., 2016) studies have been applied in tea research. However, the main limitation of linkage analysis and

association mapping is that they do not quantify the effects of the small QTLs influencing quantitative traits (Heffner et al., 2009a). Only a few major QTLs are identified while the causative rare alleles which influence important quantitative traits are overlooked (Wang et al., 2018b). The solution to the limitations associated with linkage analysis and association mapping could be addressed by genomic selection (Grattapaglia et al., 2018; Crossa et al., 2017). In GS, all marker loci are simultaneously assessed including markers with small effects that might be too small to be significant, hence improving genetic gain from selection (Jannink et al., 2010a; Meuwissen et al., 2001). The use of GS in tea breeding could lead to higher selection gains, shortened selection cycle and quicker development of new varieties, improvement of multiple traits simultaneously, and reduced genotyping and trialing costs (Steve Tanksley, Pers. com, June 2017, NSIP). Genomic selection has successfully been applied in improvement of domestic animals (VanRaden et al., 2009; Hayes et al., 2009c) and is increasingly being applied in many crops (Sverrisdóttir et al., 2017; Song et al., 2017) and trees (Grattapaglia et al., 2018; Kwong et al., 2017). However, GS has so far not been conducted in tea. In this paper, we review the conventional tea breeding process, marker assisted selection and a proposed GS framework for its implementation in tea breeding programmes.

3.2. Conventional tea breeding

Conventional tea breeding relies on phenotypic selection of promising tea varieties based on breeders' experience and the existing genetic diversity (Corley et al., 2018). It has been the main method used in selecting improved tea varieties in many countries (Mondal, 2014a; Yao et al., 2012b). All the teas cultivated worldwide originated from India and China and were introduced to other countries either directly or indirectly from these two countries (Meegahakumbura et al., 2016). Field selection and hybridization are the main approaches used in conventional tea breeding (Corley et al., 2018; Mondal, 2014b). Field selection involves identification of superior bushes in naturally existing seedling populations (Chen et al., 2007b). Promising bushes are selected and established as bi-clonal and polyclonal seed gardens (Carr, 2018). The resulting seedlings were used as parents, which have given rise to the current germplasm through crossing and selection (Meegahakumbura et al., 2018b; Carr, 2018).

Recently, modern methods of conventional tea breeding involve artificial hybridization of selected parents or with wild relatives (Meegahakumbura et al., 2018b; Sharma et al., 2010). Hybridization methods are either natural or controlled (Mondal, 2014a; Chen et al., 2007b). Natural hybridization involves using bi-clonal or polyclonal seed baries that are planted in an isolated field, allowed to flower and cross pollinate naturally (Corley et al., 2018; Mondal, 2014b). The F1 seeds from these baries are harvested, planted and the superior seedlings are selected for important traits such as yield, quality and drought tolerance. Promising seedlings are multiplied and exposed to multilocational trials to test for genotype x environment effects (Kamunya et al., 2010), and the stable varieties are multiplied and released for commercial planting (Corley et al., 2018). Polyclonal seed baries involve planting several tea bushes together, and this introduces more variability among the F1 seeds. However, it is difficult to identify the parentage of offspring resulting from polyclonal seed baries, unless molecular markers are used (Hou et al., 2006; Tanaka et al., 2005b). In controlled hand pollination, two parents with known characteristics are crossed manually by hand and subsequent generations monitored and selected for progenies with desired traits (Meegahakumbura et al., 2018b; Corley et al., 2018). The main limitation of controlled hand pollination is that the success rate is very low (Ariyarathna et al., 2011), and many crosses are required, hence making it expensive and labour intensive process (Corley et al., 2018).

3.2.1. Conventional tea breeding at Unilever Tea Kenya (UTK)

The standard method of developing superior tea varieties at Unilever Tea Kenya (UTK) is as described in Figure 1. Parents with desirable traits (yield, quality, drought tolerance, pest and disease tolerance) are selected and crossed during the first year. The seeds are germinated and allowed to grow in a tea nursery for one year, before being established in the field as hand pollinated trials (HPT). Hand pollinated trials are single progeny trials and each seedling is distinct from each other. Yield data from the HPTs are recorded for 3 years for each of the single bushes separately and the best 20% of the seedlings are selected, and vegetatively propagated in the tea nursery. Mini clonal trials (MCTs) are established in a completely randomised block design, consisting of 12 plants per clone replicated twice. Evaluation for yield

and other traits in the MCT is recorded for 5 years, and desired genotypes are selected and vegetatively propagated in the nursery for one year and then established as clonal field trials (CFT). In CFTs, the clones are planted in a completely randomised block design, consisting of 16 plants per clone replicated twice at different sites to test for the effects of genotype by environment interaction. CFTs are recorded for a further 5 years and superior varieties are selected, bulked for 3 years and released for commercial planting. The time required to develop tea varieties at Unilever Tea Kenya is 16 years (Figure 2). However, it could be reduced to 12 years if varieties are required with urgency for replanting (Table 2).

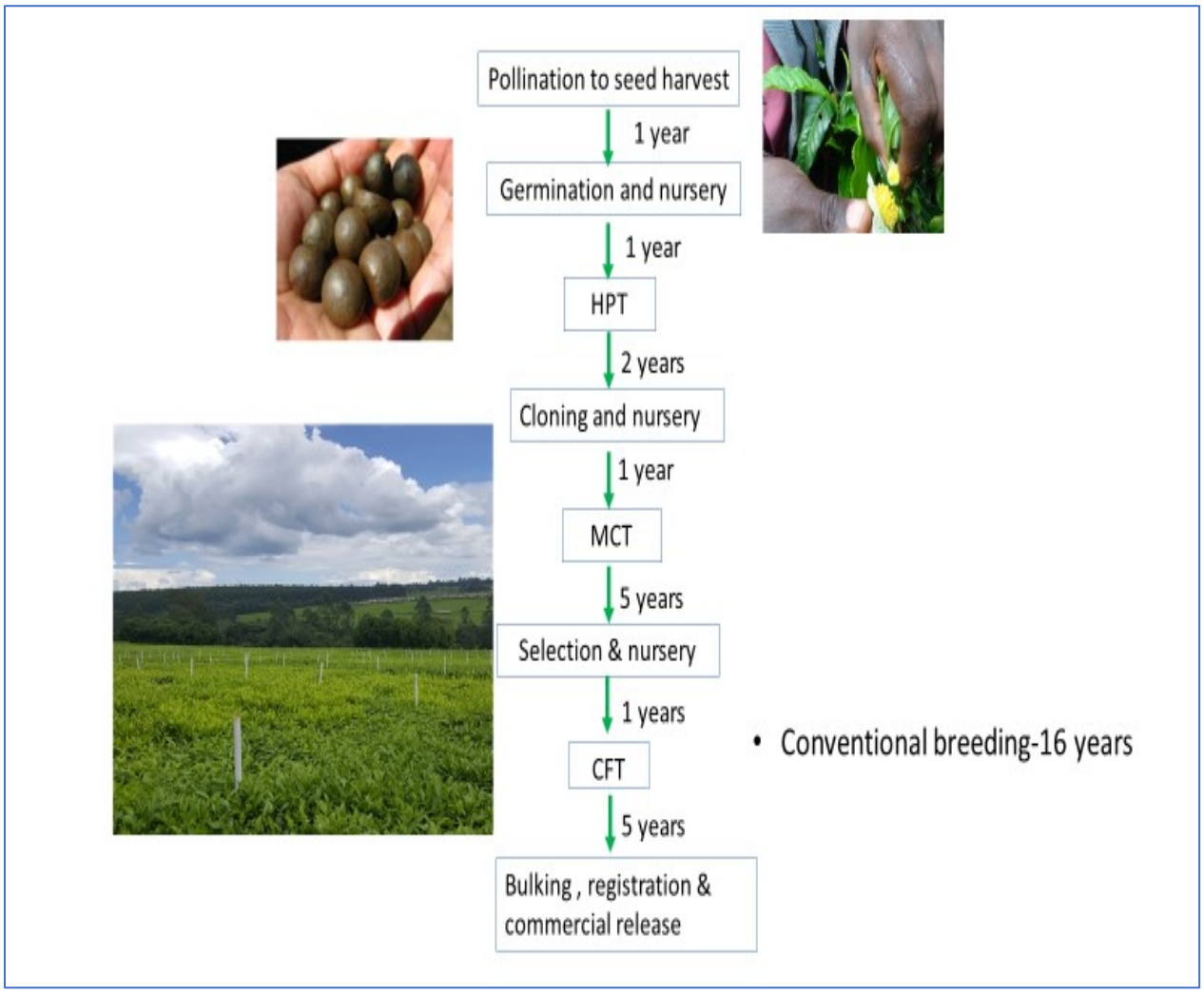

Figure 2 Conventional tea breeding at Unilever Tea Kenya Limited (HPT-hand pollinated trial, MCT-mini clonal trial, CFT- Clonal field trial).

Table 2. Accelerated tea breeding process at Unilever tea Kenya when the demand for varieties is high for replanting.

Year	Operation
0	Controlled crossing of selected clones
1	Seedlings in nursery; micro-manufacture for preliminary quality evaluation.
2	Field planting of seedlings; at least 100 per cross, if available, include standard clones
3	Yield recording of seedlings
4	Flush-shoot cuttings from the best 30% of seedlings to nursery; continue yield recording
5	Highest-yielding 20% of clones to phase 1 field trial, based on two years' records from seedlings; two replicates of 16-bush plots, plus standard clones
6-9	Yield recording of phase 1clone trial; mini-manufacture from highest yielders for quality assessment.
7-8	Cuttings from clones which yield sufficiently better than standard clones
8-9	Plant phase 2 multi-location trials with best clones; two replicates of 16-bush plots at four or five sites, representing important agroecological zones. Prune phase 1 clone trial; record trash weight and speed of recovery. Possible release for planting in same zone as phase 1 clone trial.
10-13	Yield recording of phase 2 multi-location trials Continue recording phase 1 trial for second cycle
10-11	Plant multiplication blocks of best clones
12-13	Release for commercial planting

Source: (Carr, 2018).

Conventional tea breeding is well established in the major tea growing countries and has led to development of many superior varieties (Meegahakumbura et al., 2016; Mondal, 2014b). At UTK, varieties developed through breeding are superior (better yield, quality and resistant to drought) compared to seedling genotypes (Corley et al., 2018). Historical records show that yield recorded in some of the commercial fields at Unilever Tea Kenya have ranged from 5.2 t $ha^{-1}yr^{-1}$ (seedling tea) to 9.2 t $ha^{-1}yr^{-1}$ for improved varieties (e.g. BB 35) (Table 3).

Table 3. Highest recorded yield at Unilever Tea Kenya Limited for the commonly grown commercial tea varieties compared to seedling tea at Koiwa estate.

Genotype/Cultivar	Selection criteria(s)	Maximum yield (t mt ha^{-1} yr^{-1})	Year of planting
Seedling	Unselected	5.2	1950
CLONE95	Quality	8.9	1989
CLONE98	Drought	8.3	2002
CLONE8	Yield	9.2	2014
CLONE101	Quality	6.9	1995

Source: Unilever Tea Kenya Limited.

In Eastern Africa, Kamunya et al. (2012) reported the yield range from 4.0 t ha^{-1} (CLONE106) to 6.0 t ha^{-1} (CLONE103) (Table 4). Similarly, in Southern Africa, the improved clonal teas are more superior compared to the seedling cultivars (Carr, 2018; Ellis et al., 1995). Many improved tea varieties have also been developed through breeding and clonal selection in other countries such as India (Das et al., 2012), China (Chen et al., 2007b), Sri Lanka (Gunasekare, 2012), Japan (Tanaka, 2012), Vietnam (Toan et al., 2005), Indonesia (Arfin et al., 1999) and Korea (Jeong et al., 2005).

Table 4. Some commonly grown commercial cultivars in East Africa and recorded highest yields

Genotype/Cultivar	Selection criteria(s)	Maximum yield (t mt ha^{-1} yr^{-1})	Year of release
CLONE106	Yield and anthocyanin	4.0	2011
CLONE103	Yield	6.0	2008
CLONE102	Yield	6.0	2008
TRFK 301/5	Yield and cup quality	5.9	2001
TRFK 301/4	Yield and cup quality	4.8	2001

Source: Kamunya et al. (2012).

3.2.2. Limitations of phenotypic selection in tea improvement

Conventional tea breeding has resulted in many improved varieties in many countries (Chen et al., 2007b). However, it is a labour intensive and time-consuming process (Xu et al., 2018a; Corley et al., 2018). Tea is a perennial crop with a long gestation period; it takes between 3 to 6 years for tea to grow from seedling to flowering (Mondal et al., 2004). According to Chen et al. (2007b) it can take between 22-25 years to develop improved varieties.

The selection process in conventional tea breeding leads to narrowing of the existing genetic diversity in tea, since farmers plant the few commercially released superior clones (Mukhopadhyay et al., 2016; Wachira, 2002). The controlled hand pollination process in tea has low success rates (Mondal, 2014b; Ariyarathna et al., 2011), and therefore many crosses are conducted resulting in high labour costs (Corley

et al., 2018). The tea plant also has a short flowering period of between 2–3 months in a year, and this limits the number of crosses that can be conducted in a year (Mondal, 2014b).

Genotype by environment interaction (G x E) is known to significantly impact important agronomic traits in tea such as yield (Wachira, 2002) and quality (Msomba et al., 2018; Lubang'a et al., 2015). Detecting progenies that possess favourable alleles across all loci is extremely difficult using phenotypic selection alone (Collard et al., 2005), and G × E may lead to inaccurate prediction of promising tea varieties (Kamunya et al., 2010).

3.3. Marker assisted selection (MAS)

Marker assisted selection involves the identification of genetic markers linked or close to a gene controlling a specific trait. Before marker assisted selection is utilised in plant breeding, molecular markers influencing important traits are identified through genetic mapping and validated (Collard et al., 2008; Collard et al., 2005). Genetic maps are important tools for implementing QTL analysis and MAS and are constructed using linkage mapping and association mapping techniques (Xu et al., 2018a).

3.3.1. Linkage (QTL) mapping

QTL mapping involves determination of associations between specific DNA markers and phenotypic traits of interest in a segregating bi-parental population (Collard et al., 2005). The first linkage map in tea was reported by Hackett et al. (2000), using AFLP and RAPD markers. The map had a length of 1349.7 cM and 126 markers with an average distance of 11.7 cM between loci and 15 linkage groups, corresponding to the haploid number of tea were identified. However, the map was assembled using markers from female parents only and therefore the majority of markers had unexpected segregation ratios. Later, Huang et al. (2005) constructed a linkage map with AFLP markers using female and male parents. Kamunya et al. (2010) developed a tea linkage map using RAPD, AFLP and SSR markers. The map contained 30 linkage groups and the length was 1,411.5 cM with a mean interval of 14.1 cM between loci. Twenty-three putative yield QTLs were detected in the 2 environments studied. However, these studies

used a small mapping population size comprising of only 42 offspring and therefore further validation steps were recommended before utilising the markers. The first high-density reference linkage map of tea was developed using SSR markers (Taniguchi et al., 2012). The core map had 15 linkage groups, with a total length of 1218 cM (Taniguchi et al., 2012). Ma et al. (2014) constructed a moderately saturated genetic map with 406 SSR markers from 183 genotypes that had diverse catechins content. The map had 15 linkage groups with a map length of 1,143.5 cM and 2.9 cM locus spacing. Twenty five QTLs linked to catechins were identified (Ma et al., 2014). Similarly, Tan et al. (2016b) constructed a saturated genetic map comprising of 15 linkage groups, with a length of 1226.2 cM and an average marker distance of 2.5 cM. A total of 15 QTLs were identified that were associated with timing of spring bud flush, young shoot colour, mature leaf length, mature leaf width and leaf shape index (Tan et al., 2016b). Recently a moderately saturated genetic map with a length of 1441 cM was developed using RAPD, AFLP and SSR markers (Chang et al., 2017). Japanese and Korean cultivars were used to make double pseudo-test crosses of 79 F1 mapping population.

3.3.1.1. Limitations of QTL mapping

Developing a mapping population is time-consuming, tedious and an expensive process, as it involves crossing two genetically divergent parents, differing in the traits of interest (Semagn et al., 2006b; Collard et al., 2005). However, the success rate of crossing tea is always low since many flowers abort (Ariyarathna et al., 2011). Similarly, tea is a perennial crop with a long generation time and it takes between 3 to 6 years for tea to grow from seed to flowering (Mondal, 2014b).

The available resources for conducting QTL mapping limit the size of the mapping population to be used, and this significantly affects the accuracy of detecting QTL positions and effect estimates (Heffner et al., 2009a; Schön et al., 2004). Land is a scarce, expensive and treasured resource in many countries, and this limits the number of linkage mapping experiments that can be conducted. In addition, it is expensive to employ skilled labour specifically for hand pollination and maintaining seedlings in the nursery and in

the field. Moreover, a marker validation step is required in QTL mapping, which is expensive and could limit the effectiveness of biparental populations for MAS in plant breeding (Bernardo et al., 2007).

Biparental populations used in QTL mapping are small and only capture a small portion of the total genetic variation, hence do not represent the whole breeding population (Xu, 2010b). Studies conducted in tea, have shown that the small population used in QTL mapping leads to overestimated and spurious QTL effects (Kamunya et al., 2010). Additionally, the statistical methods used in linkage analysis are not adequate for improving quantitative traits, that are controlled by many genes each with small effect (Jannink et al., 2010a; Heffner et al., 2010).

3.3.2. The future of tea breeding: key traits and marker-traits analysis

Association mapping is a population-based method used to detect and map QTLs based on significant association of molecular markers and phenotypic traits (Gupta et al., 2005). It exploits the historical linkage disequilibrium to identify marker-trait relationships within a natural non-biparental population (Pasam et al., 2012), resulting in higher mapping resolution (Yu et al., 2011; Stich et al., 2010). Association mapping was first applied in human genetics to study associations between QTLs and diseases (Hindorff, 2009). More recently, association mapping has been applied in many crops (Soto-Cerda et al., 2012b; Hall et al., 2010; Sorrells et al., 2009). However, very few association mapping studies have been conducted in tea. SNPs associated with caffeine content were identified, validated and recommended for use in marker-assisted selection for tea quality improvement (Jin et al., 2016). Similarly, Su et al. (2016) studied association between quality traits in tea and markers in China, detecting 2211 pairs of loci of which 259 had significant linkage disequilibrium with the traits ($D' > 0.5$). 19 EST-SSR markers were found to be significantly associated with polyphenols, total free amino acids and caffeine. Findings from this work can be applied directly in improvement of tea quality (Su et al., 2016).

3.3.2.1. Advantages of association mapping

Association mapping is conducted on the existing natural population, which is readily available and therefore identification of QTLs does not involve developing a bi-parental population which is expensive and time-consuming. Similarly, there are higher levels of genetic recombination within an association mapping population, which leads to a higher resolution compared to the bi-parental population applied in QTL mapping (Heffner et al., 2009b). Additionally, many alleles can be evaluated simultaneously for one locus, and this ensures effective gene tagging and accurate identification of markers for use in marker assisted selection. Association mapping is also more cost effective compared to QTL mapping since many traits can be investigated using the same panel of accessions and genotypic data without the need for an expensive validation procedure (Breseghello et al., 2006).

3.3.2.2. Limitations of association studies

Although association mapping leads to a greater resolution compared to QTL mapping, it often leads to overestimation of effects due to the arbitrary significance thresholds set, which sometimes leads to poor trait predictions (Heffner et al., 2009a; Schön et al., 2004). Similarly, the population structure and kinship within an association mapping population may lead to spurious associations, hence predicting false positive marker-trait association (Zhao et al., 2007b). In addition, the detection of a marker-trait association is influenced by allele frequency distribution within a population and therefore, it is difficult to identify a causative rare allele (Myles et al., 2009). This may lead to biased estimation of effect and poor prediction of performance of individuals (Rafalski, 2002). Even when a potentially useful allele is identified, it may be present in a poor quality or yield individual, meaning significant backcrossing may be needed to allow the allele to be introgressed into elite germplasm – a very long process in tea.

3.4. The future of tea breeding: a new framework based on genomic selection

GS is a type of MAS that simultaneously estimates all loci, haplotype and marker effects across the entire genome to produce an estimate of the GEBVs of individuals (Heffner et al., 2009a; Meuwissen et al.,

2001). A training population with phenotypes and genotypes are used to estimate GEBVs using prediction models, based on the association between genotypes and phenotypes. The model captures total additive genetic variance across the entire genome to estimate GEBVs among the selection candidates based on the sum of all marker effects (Lorenz et al., 2011a; Heffner et al., 2010; Heffner et al., 2009b). In GS, markers are treated as random effects and all marker effects are estimated simultaneously in a single GS model, with the use of part of the dataset as a validation of the model, predicting phenotypic values from genotypes and then testing against actual recorded phenotypic values to give a correlation score (Lorenz et al., 2011b; 2011a; Heffner et al., 2011; Hayes et al., 2009b). Subsequently, GEBVs are estimated for a common or closely related breeding population using only genotypic data, for use in selecting good breeding candidates (Heffner et al., 2009b; Goddard et al., 2007; Meuwissen et al., 2001), followed by selection of the offspring and recrossing without gathering phenotypic data. GS has the potential of shortening breeding cycles of crops, increasing selection intensity and improving the accuracy of selection (Grattapaglia et al., 2018; Crossa et al., 2017).

In GS, the number of molecular markers are usually greater than the number of phenotypic measurements, implying that there are more predictor variables compared to phenotypic observations, hence creating a “large p and small n problem” (Lorenz et al., 2011a). The degrees of freedom are not adequate for estimating all predictor effects simultaneously using the ordinary least squares approach (Lorenz et al., 2011a). Even if there were sufficient degrees of freedom, a high degree of multicollinearity among markers may occur leading to an overfitted model that exaggerates minor fluctuations, and consequently reduces prediction accuracy (Wang et al., 2018b; Grattapaglia et al., 2018; Lorenz et al., 2011a; Jannink et al., 2010b). A variety of statistical models have been developed to solve the problem of having large numbers of markers and few phenotypes, by estimating the effect of each marker (SNP effects) and accurately predicting traits of economic importance such as yield, quality, resistance to biotic and abiotic stresses. These models differ in the assumptions on variance of marker effect and the type of gene action (Lorenz et al., 2011b). The main statistical models used in genomic selection include ridge regression best linear unbiased predictor (RR-BLUP), genomic best linear unbiased predictor (G-BLUP), Bayesian

models (BayesA, BayesB, BayesC, Bayes LASSO) and machine learning (Wang et al., 2018b; Lorenz et al., 2011a). The accuracy of genomic selection is affected by linkage disequilibrium (LD) between markers and QTL, marker type and density, number of QTLs affecting the trait, heritability, the model, size and structure of the population, genetic relationship between the training and the breeding population and genetic architecture of traits (Wang et al., 2018b; Zhong et al., 2009b).

3.4.1. Implementation of GS in other crops

GS has successfully been tested in many crops with success e.g. wheat (Juliana et al., 2017; Bassi et al., 2016), cassava (Wolfe et al., 2017), maize (Cerrudo et al., 2018; Crossa et al., 2013), oil palm (Kwong et al., 2017), potatoes (Stich et al., 2018), forest trees (Grattapaglia, 2014) and rice (Spindel et al., 2018). In oil palm, crosses are made, and individuals selected based on markers alone hence limiting the use of progeny tests to the training of the GS model (Cros et al., 2014). In forest trees, high prediction accuracies have been obtained which are greater than or equal to those obtainable by pedigree-based phenotypic selection (Grattapaglia, 2014; Denis et al., 2013). In rice breeding, GS has been implemented as shown in Figure 3. Crosses are made and about 20,000 progenies are developed. GEBVs are estimated from the F3 population using a suitable GS model, and selections are made until the F8 generations (Spindel et al., 2018).

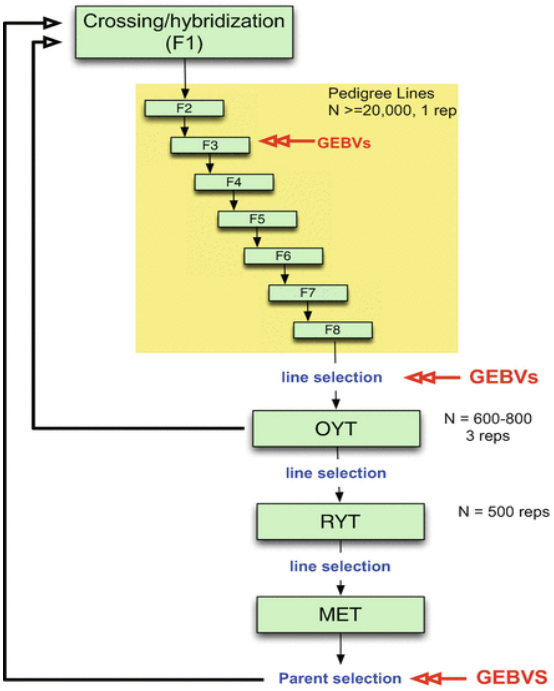

Figure 3. Rice breeding pipeline that integrates GS. Whereby; OYT (observational yield trial), RYT (replicated yield trials), MET (multi-environment trials) (Spindel et al., 2018).

3.4.2. Genomic selection in tea breeding

The major challenge facing all tea breeding programmes is the long generation interval, as it takes between 3 to 6 years for tea to grow from seedling to flowering (Mondal, 2014b). The breeder's equation is ($\Delta G = ir\sigma A/L$), where i is the intensity of selection, r is the selection accuracy (heritability), σA is the additive genetic variance and L is the breeding cycle interval. Adopting GS in tea breeding can potentially increase i since large numbers of progenies at the nursery stage can be predicted using only molecular marker data. Moreover, GS uses realized genomic relationship matrix that increases the accuracy of estimating σA and GEBVs (Grattapaglia et al., 2018; El-Dien et al., 2018).

The greatest impact of GS on the genetic gain in tea breeding will be from decreasing (L). The period required to conventionally develop an improved tea variety is between 22-25 years (Chen et al., 2007b), which includes crossing, field evaluation of progeny crosses and performing selections, and propagation of selected improved materials vegetatively. GS promises to significantly reduce the length

required to develop improved varieties and hence increase genetic gain per time unit by bypassing some of the field-testing steps. By adopting GS in tea breeding, the phenotypes of the breeding population could be predicted at very early stages when the seedlings are still in the nursery. In addition, implementation of GS in tea breeding will allow simultaneous and early selection for multiple traits in many individuals, which is not possible in conventional tea breeding. Similarly, the use of accurate GS models and reduced genotyping cost will lead to improvement in the efficiency of the tea breeding programme.

3.4.3. Genomic selection: case for Unilever

At the time of writing, there have not been any trials of GS in tea breeding. Adoption of GS in tea breeding is expected to improve the accuracy and reduce the time and cost of developing improved varieties. We have proposed a scheme for implementing GS in the UTK breeding programme (Figure 4). The scheme takes advantage of the increase in genetic gains generated by developing an improved tea variety in 7 years.

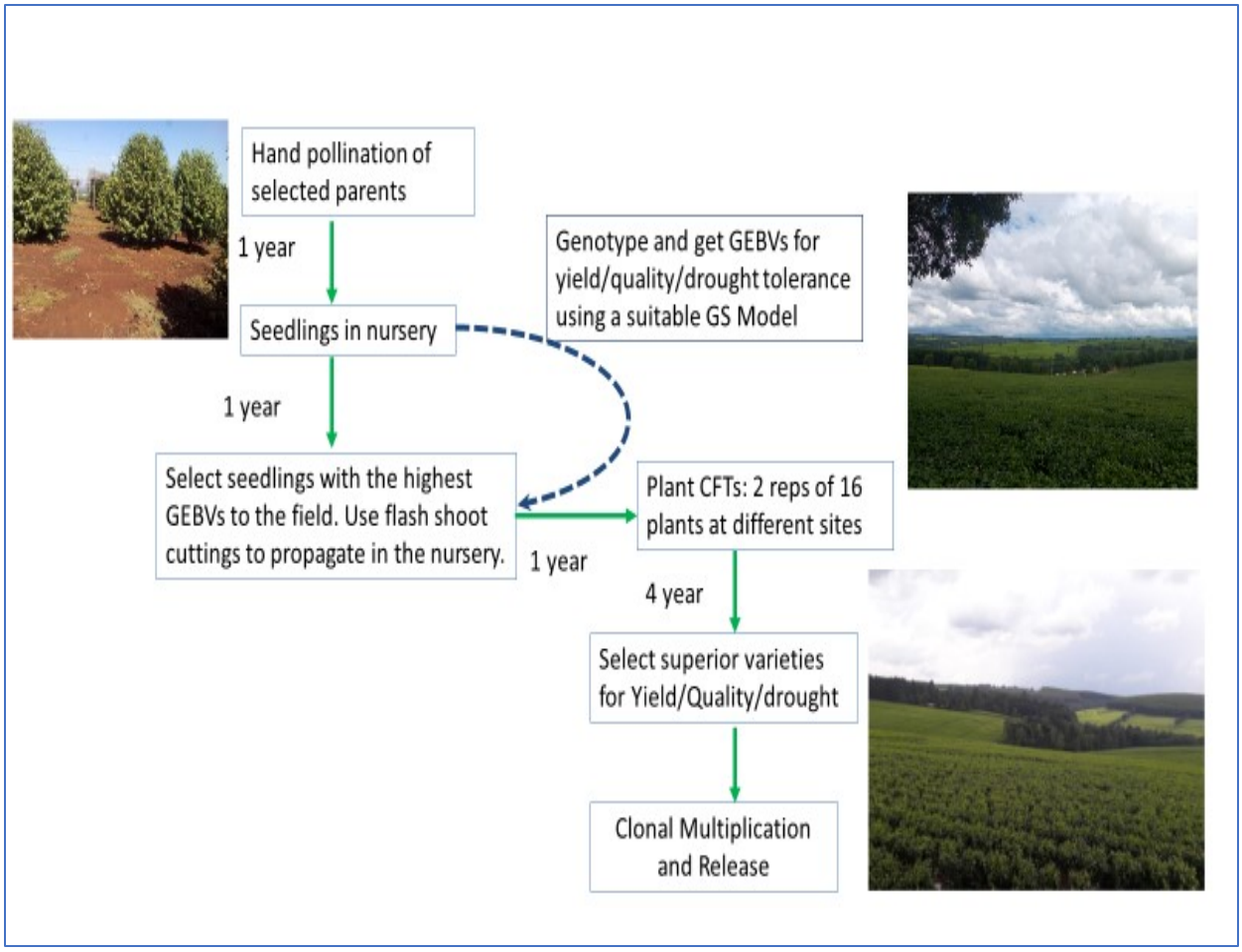

Figure 4. Proposed GS implementation method in tea breeding programmes. Whereby; CFT- Clonal field trials, GS (Genomic selection), GEBVs (Genome estimated breeding values).

Parents in the germplasm garden with traits of interest are selected and crossed manually (hand pollination). Hand pollination and seed maturation process takes one year. The mature seeds are pre-germinated in the seed germination room and planted in the nursery for one year. While in the nursery, DNA is extracted from the seedlings, genotyped and GEBVs estimated. The seedlings with the highest GEBVs are selected, flush shoots obtained from them and propagated in the nursery. The propagated shoots are allowed to grow for one year in the nursery and planted in the field as CFTs at different sites to test for genotype by environment interaction (G x E). Yield, quality and drought tolerance data is collected from

the CFTs for 4 years. Clones that are stable and superior across the sites and at specific sites are selected, released and bulked for commercial planting after 4 years of data collection in the CFTs.

3.4.4. Implication of GS in tea breeding

The best way of gaining more from GS in tea breeding is by applying GS early at the nursery stage. Conventional tea breeding has many bottlenecks namely; 1. perennial nature of tea , 2. long gestation periods, 3. high inbreeding depression, 4. self-incompatibility, 5. lack of distinct selection criteria, 6. low success of hand pollination, 7. short flowering season, 8. long duration of seed maturation (Mondal, 2014b). Implementing GS in tea breeding could solve some of these challenges. The general process of applying the proposed GS scheme involves making crosses, selecting the best seedlings based on GEBVs, field evaluation for yield, quality and drought tolerance, testing for G x E effects and selecting superior tea varieties for commercial planting. GS can be used in such schemes as early as the second year of crossing hence reducing breeding cycle time drastically. When compared to phenotypic selection, this scheme can improve genetic gain per unit time. For instance, by implementing GS using the proposed scheme whereby the HPTs (Hand pollinated trials) and MCTs (Mini clonal trials) are bypassed, tea breeders are able to avoid the additional lengthy and expensive phenotyping steps from 16 years (Figure 2) to 7 years (Figure 4) thus saving time and money. Additionally, GS could be applied to individuals that have not been phenotyped but are closely related to the training population, hence increasing the selection intensity, since less effort and time would be required for genotyping the selection candidates compared to progeny testing.

3.5. Conclusion

Conventional tea breeding has been the main method of developing improved varieties in tea and has served tea breeders well but there are bottlenecks, which molecular breeding could help. The literature shows limited use of MAS especially derived from linkage mapping and association mapping in tea improvement. GS has been successful in commercial plantation crops e.g. oil palm, bananas and forest trees, and could be adopted in tea breeding. GS has so far not been implemented in tea and the prospects of

applying GS to tea breeding are very promising. We have proposed one GS approach which takes 7 years to develop an improved tea variety. Adopting this approach in tea breeding would not only improve breeding efficiency (save time) but would also reduce the cost and increase the chance of developing superior varieties.

References

Arfin, T., Semangun, P. (1999). Tea industry in Indonesia. *Aravali Book International Pvt Ltd, New Delhi*, 65–72.

Ariyarathna, H. A. C. K., Gunasekare, M. T. K., Kottawa-Arachchige, J. D., Paskarathevan, R., Ranaweera, K. K., Ratnayake, M., Kumara, J. B. D. A. P. (2011). Morpho-physiological and phenological attributes of reproductive biology of tea (Camellia sinensis (L.) O. Kuntze) in Sri Lanka. *Euphytica, 181*(2), 203-215. doi:10.1007/s10681-011-0399-9

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp .). *Plant Science, 242*, 23-36. doi:10.1016/j.plantsci.2015.08.021

Bernardo, R., Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Science, 47*(3), 1082-1090.

Breseghello, F., Sorrells, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Science, 46*(3), 1323-1330.

Carr, M. K. V. (2018). *Advances in Tea Agronomy*: Cambridge University Press.

Cerrudo, D., Cao, S., Yuan, Y., Martinez, C., Suarez, E. A., Babu, R., Zhang, X., Trachsel, S. (2018). Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments. *Frontiers in Plant Science, 9*. doi:10.3389/fpls.2018.00366

Chang, H. T. (1981). Thea—a section of beveragial tea trees of the genus Camellia. *Acta Sci Nat Univ Sunyats, 1*, 87–99.

Chang, Y., Oh, E. U., Lee, M. S., Kim, H. B., Moon, D.-G., Song, K. J. (2017). Construction of a genetic linkage map based on RAPD, AFLP, and SSR markers for tea plant (Camellia sinensis). *Euphytica, 213*(8), 190.

Chen, L., Zhou, Z.-X., Yang, Y.-J. (2007). Genetic improvement and breeding of tea plant (Camellia sinensis) in China: from individual selection to hybridization and molecular breeding. *Euphytica, 154*(1-2), 239-248.

Chen, Z.-M., Chen, L. (2012). Delicious and healthy tea: an overview. In *Global tea breeding* (pp. 1-11): Springer.

Collard, B. C. Y., Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1491), 557-572. doi:10.1098/rstb.2007.2170

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196. doi:10.1007/s10681-005-1681-5

Corley, R. H. V., Tuwei, G. (2018). The Well-Bred Tea Bush. *Cambridge University Press.In: M. Carr (Author), Advances in Tea Agronomy*, 106-136.

Cros, D., Denis, M., Sánchez, L., Cochard, B., Flori, A., Durand-Gasselin, T., Nouy, B., Omoré, A., Pomiès, V., Riou, V., Suryana, E., Bouvet, J.-M. (2014). Genomic selection prediction accuracy in a perennial crop: case study of oil palm (Elaeis guineensis Jacq.). *Theoretical and Applied Genetics, 128*(3), 397-410. doi:10.1007/s00122-014-2439-z

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., De Los Campos, G., Burgueño, J., Windhausen, V. S., Buckler, E. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3: Genes, Genomes, Genetics*, g3-113.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science, 22*(11), 961-975.

Das, S. C., Das, S., , Hazarika, M. (2012). Breeding of the Tea Plant (Camellia sinensis) in India. In N. M. (Ed.), *Global Tea Breeding. Advanced Topics in Science and Technology in China.* (pp. 69-124.). Berlin, Heidelberg, : Springer.

Denis, M., Bouvet, J.-M. (2013). Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics & Genomes, 9*(1), 37-51.

Devarumath, R., Nandy, S., Rani, V., Marimuthu, S., Muraleedharan, N., Raina, S. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing Camellia sinensis (China type) and C. assamica ssp. assamica (Assam-India type). *Plant Cell Reports, 21*(2), 166-173.

El-Dien, O. G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., El-Kassaby, Y. A. (2018). Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (Picea glauca x engelmannii). *Molecular Breeding, 38*(3), 26.

Ellis, R., Nyirenda, H. E. (1995). A successful plant improvement programme on tea (Camellia sinensis). *Experimental Agriculture, 31*(3), 307-323.

Goddard, M. E., Hayes, B. J. (2007). Genomic selection. *Journal of animal breeding and genetics, 124*(6), 323-330.

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S. F., Tan, B., Isik, F., Ratcliffe, B., El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science, 9*.

Grattapaglia, D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. In *Genomics of plant genetic resources* (pp. 651-682): Springer.

Gunasekare, M. T. K. (2012). Tea plant (Camellia sinensis) breeding in Sri Lanka. In *Global Tea Breeding* (pp. 125-176): Springer.

Gupta, P. K., Rustgi, S., Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology, 57*(4), 461-485.

Hackett, C. A., Wachira, F. N., Paul, S., Powell, W., Waugh, R. (2000). Construction of a genetic linkage map for Camellia sinensis (tea). *Heredity, 85*(4), 346.

Hall, D., Tegström, C., Ingvarsson, P. K. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in functional genomics, 9*(2), 157-165.

Hashimoto, M., Takasi, S. ((1978). Morphological studies on the origin of the tea plant V, a proposal of one place of origin by cluster analysis. *J Trop Agric, 21*, 93–101.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science, 92*(2), 433-443. doi:10.3168/jds.2008-1646

Hayes, B. J., Visscher, P. M., Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research, 91*(1), 47-60.

Heffner, E. L., Sorrells, M. E., Jannink, J.-L. (2009a). Genomic selection for crop improvement. *Crop Science, 49*(1), 1-12.

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Science, 50*(5), 1681-1690.

Heffner, E. L., Sorrells, M. E., Jannink, J.-L. (2009b). Genomic Selection for Crop Improvement. *Crop Science, 49*(1). doi:10.2135/cropsci2008.08.0512

Heffner, E. L., Jannink, J.-L., Iwata, H., Souza, E., Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science, 51*(6), 2597-2606.

Hindorff, L. A. (2009). A catalog of published genome-wide association studies. *http://www. genome. gov/26525384*.

Hou, Y. J., He, Q., Liang, G. L., Li, P. W., Peng, P., Deng, M. (2006). ISSR analysis of the hybrid of the descendants of tea camellias. *Journal of Southwest Agricultural University (Natural Science), 28*(2), 267-270.

Huang, J. A., Li, J. X., Huang, Y. H., Luo, J. W., Gong, Z. H., Liu, Z. H. (2005). Construction of AFLP molecular markers linkage map in tea plant. *J Tea Sci, 25*(1), 7-15.

Jannink, J.-L., Lorenz, A. J., Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics, 9*(2), 166-177.

Jannink, J. L., Lorenz, A. J., Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics, 9*(2), 166-177. doi:10.1093/bfgp/elq001

Jeong, B., Song, Y., Moon, Y., Han, S., Bang, J., Kim, J., Kim, J., Park, Y. (2005, Nov 11–15.). *Tea tree breeding plans for the tea Industry in Korea.* Paper presented at the International Tea Symposium-2005, China.

Jin, J.-Q., Yao, M.-Z., Ma, C.-L., Ma, J.-Q., Chen, L. (2016). Association mapping of caffeine content with TCS1 in tea plant and its related species. *Plant Physiology and Biochemistry, 105*, 251-259.

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., Bhavani, S., Rutkoski, J. E., Poland, J. A., Bergstrom, G. C. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theoretical and Applied Genetics, 130*(7), 1415-1430.

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Muoki, R. C., Sharma, R. K. (2012). Tea improvement in Kenya. In *Global Tea Breeding* (pp. 177-226): Springer.

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Korir, R., Sharma, V., Kumar, R., Bhardwaj, P., Chalo, R., Ahuja, P. S., Sharma, R. K. (2010). Genomic mapping and testing for quantitative trait loci in tea (Camellia sinensis (L.) O. Kuntze). *Tree Genetics & Genomes, 6*(6), 915-929. doi:10.1007/s11295-010-0301-2

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 Genetics & Genomes, 14*(1). doi:10.1007/s11295-017-1219-8

Kondo, K. (1977). Chromosome numbers in the genus Camellia. *Biotropica*, 86-94.

Kwong, Q. B., Ong, A. L., Teh, C. K., Chew, F. T., Tammi, M., Mayes, S., Kulaveerasingam, H., Yeoh, S. H., Harikrishna, J. A., Appleton, D. R. (2017). Genomic Selection in Commercial Perennial Crops: Applicability and Improvement in Oil Palm (Elaeis guineensis Jacq.). *Sci Rep, 7*(1), 2872. doi:10.1038/s41598-017-02602-6

Leng, P.-f., Lübberstedt, T., Xu, M.-l. (2017). Genomics-assisted breeding – A revolutionary strategy for crop improvement. *Journal of Integrative Agriculture, 16*(12), 2674-2685. doi:10.1016/s2095-3119(17)61813-6

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., Smith, K. P., Sorrells, M. E., Jannink, J.-L. (2011a). Genomic selection in plant breeding: knowledge and prospects. In *Advances in agronomy* (Vol. 110, pp. 77-123): Elsevier.

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., Smith, K. P., Sorrells, M. E., Jannink, J.-L. (2011b). Genomic Selection in Plant Breeding. In (pp. 77-123).

Lubang'a, N. M., Kamunya, S. M., Kiplagat, O., Wanyoko, J. K., Chalo, R. M. (2015). Effects of genotype by environment interaction on quality traits in Kenyan tea (Camellia sinensis)(L.) O. Kuntze. *Tea, 36*(1), 24-39.

Ma, J.-Q., Yao, M.-Z., Ma, C.-L., Wang, X.-C., Jin, J.-Q., Wang, X.-M., Chen, L. (2014). Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (Camellia sinensis). *PLoS One, 9*(3), e93131.

Meegahakumbura, M. K., Wambulwa, M. C., Li, M.-M., Thapa, K. K., Sun, Y.-S., Möller, M., Xu, J.-C., Yang, J.-B., Liu, J., Liu, B.-Y. (2018). Domestication Origin and Breeding History of the Tea Plant (Camellia sinensis) in China and India Based on Nuclear Microsatellites and cpDNA Sequence Data. *Frontiers in Plant Science, 8*, 2270.

Meegahakumbura, M. K., Wambulwa, M. C., Thapa, K. K., Li, M. M., Möller, M., Xu, J. C., Yang, J. B., Liu, B. Y., Ranjitkar, S., Liu, J. (2016). Indications for three independent domestication events for the tea plant (Camellia sinensis (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PLoS One, 11*(5), e0155369.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.

Mondal, T. K. (2014a). Genetics and Breeding. In *Breeding and Biotechnology of Tea and its Wild Species* (pp. 9-34): Springer.

Mondal, T. K. (2014b). *Breeding and biotechnology of tea and its wild species*: Springer Science & Business Media.

Mondal, T. K., Bhattacharya, A., Laxmikumaran, M., Ahuja, P. S. (2004). Recent advances of tea (Camellia sinensis) biotechnology. *Plant Cell, Tissue and Organ Culture, 76*(3), 195-254.

Msomba, S. W., Kamunya, S. M., Rweyemamu, C. L., Reuben, S. O. W. M. (2018). EFFECT OF STABILITY AND ADAPTABILITY ON TEA QUALITY IN DIVERSE GROWING ENVIRONMENTS OF TANZANIA.

Mukhopadhyay, M., Mondal, T. K., Chand, P. K. (2016). Biotechnological advances in tea (Camellia sinensis [L.] O. Kuntze): a review. *Plant Cell Reports, 35*(2), 255-287.

Muoki, R. C., Wachira, F. N., Pathak, R. S., Kamunya, S. M. (2007). Assessment of the mating system of Camellia sinensis in biclonal seed orchards based on PCR markers. *The Journal of Horticultural Science and Biotechnology, 82*(5), 733-738.

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., Buckler, E. S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell, 21*(8), 2194-2202.

Pasam, R. K., Sharma, R., Malosetti, M., van Eeuwijk, F. A., Haseneyer, G., Kilian, B., Graner, A. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biology, 12*(1), 16.

Rafalski, J. A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science, 162*(3), 329-333.

Schön, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics, 167*(1), 485-498.

Semagn, K., Bjørnstad, Å., Ndjiondjop, M. N. (2006). Principles, requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology, 5*(25).

Sharma, R. K., Negi, M. S., Sharma, S., Bhardwaj, P., Kumar, R., Bhattachrya, E., Tripathi, S. B., Vijayan, D., Baruah, A. R., Das, S. C. (2010). AFLP-based genetic diversity assessment of commercially important tea germplasm in India. *Biochemical genetics, 48*(7-8), 549-564.

Song, J., Carver, B. F., Powers, C., Yan, L., Klápště, J., El-Kassaby, Y. A., Chen, C. (2017). Practical application of genomic selection in a doubled-haploid winter wheat breeding program. *Molecular Breeding, 37*(10). doi:10.1007/s11032-017-0715-8

Sorrells, M. E., Yu, J. (2009). Linkage disequilibrium and association mapping in the Triticeae. In *Genetics and genomics of the Triticeae* (pp. 655-683): Springer.

Soto-Cerda, B. J., Cloutier, S. (2012). Association mapping in plant genomes. In *Genetic diversity in plants*: InTech.

Spindel, J., Iwata, H. (2018). Genomic Selection in Rice Breeding. In T. Sasaki M. Ashikari (Eds.), *Rice Genomics, Genetics and Breeding* (pp. 473-496). Singapore: Springer Singapore.

Stich, B., Melchinger, A. E. (2010). An introduction to association mapping in plants. *CAB Rev, 5*, 1-9.

Stich, B., Van Inghelandt, D. (2018). Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. *Frontiers in Plant Science, 9*, 159-159. doi:10.3389/fpls.2018.00159

Su, H., Liu, J., He, W., Wen, B., Li, M., Feng, J. (2016). Association analysis of traits related to tea (Camellia sinensis) quality with EST-SSRs in Southern Henan area. *Journal of Agricultural Biotechnology, 24*(9), 1328-1336.

Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., Nielsen, K. L. (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics, 130*(10), 2091-2108.

Tan, L.-Q., Wang, L.-Y., Xu, L.-Y., Wu, L.-Y., Peng, M., Zhang, C.-C., Wei, K., Bai, P.-X., Li, H.-L., Cheng, H., Qi, G.-N. (2016). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (Camellia sinensis). *Tree Genetics & Genomes, 12*(3), 52. doi:10.1007/s11295-016-1008-9

Tanaka, J. (2012). Japanese Tea Breeding History and the Future Perspective. In *Global Tea Breeding: Achievements, Challenges and Perspectives* (pp. 227-239). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tanaka, T., Mizutani, T., Shibata, M., Tanikawa, N., , Parks, C. R. (2005). Cytogenetic studies on the origin of Camellia × vernalis. V. Estimation of the seed parent of C. × vernalis that evolved about 400 years ago by cpDNA analysis. . *J Jap Soc Hort Sci., 74*, 464–468.

Taniguchi, F., Furukawa, K., Ota-Metoku, S., Yamaguchi, N., Ujihara, T., Kono, I., Fukuoka, H., Tanaka, J. (2012). Construction of a high-density reference linkage map of tea (Camellia sinensis). *Breeding Science, 62*(3), 263-273. doi:10.1270/jsbbs.62.263

Toan, N. V., Tao, N. V. (2005, Nov 11–15). *Tea breeding selection by hybridization method in Vietnam.* Paper presented at the International Tea Symposium 2005, Hangzhou, China.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., Schenkel, F. S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science, 92*(1), 16-24. doi:https://doi.org/10.3168/jds.2008-1514

Varshney, R. K., Terauchi, R., McCouch, S. R. (2014). Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLOS Biology, 12*(6), e1001883. doi:10.1371/journal.pbio.1001883

Wachira, F. N., Kiplangat, J. K. (1991). Newly identified Kenyan Polyploid Tea Strains. . *Tea, 12*(1), 10-13.

Wachira, F. N. (2002). Genetic diversity and characterisation of Kenyan tea germplasm. *A tea Genetic Diversity (TGD) project. TGD final project document, Kericho, Kenya*.

Wachira, F. N., Kamunya, S. K. (2005). Pseudo-self-incompatibility in some tea clones (Camellia sinensis (L.) O. Kuntze). *The Journal of Horticultural Science and Biotechnology, 80*(6), 716-720. doi:10.1080/14620316.2005.11512004

Wang, X., Xu, Y., Hu, Z., Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal, 6*(4), 330-340. doi:https://doi.org/10.1016/j.cj.2018.03.001

Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., Xia, E., Lu, Y., Tai, Y., She, G. (2018). Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proceedings of the National Academy of Sciences*, 201719622.

Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., Lozano, R., Okeke, U. G., Ozimati, A. A., Williams, E., Egesi, C., Kawuki, R. S., Kulakow, P., Rabbi, I. Y., Jannink, J.-L. (2017). Prospects for Genomic Selection in Cassava Breeding. *The Plant Genome, 10*(3). doi:10.3835/plantgenome2017.03.0015

Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., Zhang, Y., Liu, Y., Zhu, T., Li, W., Huang, H., Tong, Y., Nan, H., Shi, C., Shi, C., Jiang, J.-J., Mao, S.-Y., Jiao, J.-Y., Zhang, D., Zhao, Y., Zhao, Y.-J., Zhang, L.-P., Liu, Y.-L., Liu, B.-Y., Yu, Y., Shao, S.-F., Ni, D.-J., Eichler, E. E., Gao, L.-Z. (2017). The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Molecular Plant, 10*(6), 866-877. doi:10.1016/j.molp.2017.04.002

Xu, L. Y., Wang, L. Y., Wei, K., Tan, L. Q., Su, J. J., Cheng, H. (2018). High-density SNP linkage map construction and QTL mapping for flavonoid-related traits in a tea plant (Camellia sinensis) using 2b-RAD sequencing. *BMC Genomics, 19*(1), 955. doi:10.1186/s12864-018-5291-8

Xu, Y. Q. (2010). Molecular plant breeding. *CABI, Wallingford, UK*, 736

Yao, M.-Z., Chen, L. (2012a). Tea Germplasm and Breeding in China. In *Global Tea Breeding: Achievements, Challenges and Perspectives* (pp. 13-68). Berlin, Heidelberg: Springer Berlin Heidelberg.

Yao, M.-Z., Chen, L. (2012b). Tea germplasm and breeding in China. In *Global Tea Breeding* (pp. 13-68): Springer.

Ye, Y., Yan, J., Cui, J., Mao, S., Li, M., Liao, X., Tong, H. (2018). Dynamic changes in amino acids, catechins, caffeine and gallic acid in green tea during withering. *Journal of Food Composition and Analysis, 66*, 98-108. doi:10.1016/j.jfca.2017.12.008

Yu, X., Bai, G., Luo, N., Chen, Z., Liu, S., Liu, J., Warnke, S. E., Jiang, Y. (2011). Association of simple sequence repeat (SSR) markers with submergence tolerance in diverse populations of perennial ryegrass. *Plant Science, 180*(2), 391-398. doi:https://doi.org/10.1016/j.plantsci.2010.10.013

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M. (2007). An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS genetics, 3*(1), e4. doi:10.1371/journal.pgen.0030004

Zhong, S., Dekkers, J. C. M., Fernando, R. L., Jannink, J. L. (2009). Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics, 182*(1), 355-364. doi:10.1534/genetics.108.098277

Advances in techniques and methods for assessing tea (*Camellia sinensis* (L.) O. Kuntze) quality: A review of recent development

Lubanga M. Nelson[b,c*], De-Silva Jacquie[d], Massawe Festo[b], Mayes Sean[a]

*nelsonlubanga@gmail.com

[a]School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK.

[b]School of biosciences, University of Nottingham, Malaysia Campus Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[c]Unilever Tea Kenya Ltd., Nakuru Road, Kericho, 20-20200, Kenya.

[d]Unilever Research & Development, Colworth Science Park, Sharnbrook, Bedford MK441LQ, UK.

Abstract

Manufactured tea quality for *Camellia sinensis* (L.) O. Kuntze is an important factor as it influences the consumer acceptability and the price of manufactured tea during international trading at the tea auction. In the recent years, there has been an oversupply of tea in the world, and high-quality tea is given more preference by the buyers. Consequently, most tea companies are currently investing in research focusing on breeding and manufacturing of high-quality tea. However, evaluating tea quality is currently the most challenging process. At present, tea quality is evaluated commercially using sensory evaluation method based on experienced tea tasters, which can be subjective. This makes tea quality evaluation a challenging process. It is important to evaluate the quality of tea accurately as it will benefit the consumers and tea companies, as well as breeding programmes. Physical and chemical methods that use spectroscopy, chromatography and capillary electrophoresis have also been developed and used to accurately evaluate tea quality. Recently, bionics techniques such as electronic nose, electronic tongue, computer vision and multi-sensors been developed and are promising tools in tea evaluation. Although the physical and chemical methods are accurate and reliable, they are expensive, time consuming and require skilled personnel. Bionics techniques are rapid, cost effective, non-destructive and easy to use, and have a potential to

complement or replace sensory evaluation. However, more research is needed to optimize bionics techniques and facilitate their use and adoption in quality evaluation by the tea industry.

Key words; Tea quality, sensory evaluation, analytical techniques, bionics

4.1. Introduction

Tea (*Camellia sinensis* (L.) O. Kuntze) quality is the main factor influencing the price of tea (Gallaher et al., 2006) and is defined by the colour and flavour of made tea (Zheng et al., 2016a; Ho et al., 2015; Wang et al., 2010). Fresh green leaf contains many biochemical components, which are grouped into aroma compounds, flavonoids, alkaloids, amino acids, carbohydrates, lipids, vitamins and pigments (Abdel-Rahman et al., 2011; Caffin et al., 2004; Harbowy et al., 1997). However, only aroma compounds, polyphenols, caffeine and amino acids have the greatest influence on the quality of black tea (Tu et al., 2018; Koch et al., 2018; Magagna et al., 2017; Zheng et al., 2016a; Ho et al., 2015; Owuor et al., 2007). The polyphenols in fresh harvested tea leaves are oxidised into theaflavins and thearubigins during black tea processing, catalysed by the polyphenol oxidase enzyme (Li et al., 2013; Kumar et al., 2013; Friedman et al., 2005). Theaflavins contribute to the taste and colour of black tea (Ghosh et al., 2012; Owuor et al., 1998; Owuor et al., 1994), while thearubigins contribute to the mouth feel (thickness) and colour of the tea (John et al., 2014; Owuor et al., 2004).

The current methods used to evaluate tea quality are either subjective or objective (Chen et al., 2018; Liang et al., 2003). Sensory evaluation using professional tasters has been used for many years to evaluate, describe, grade and determine the price of tea (Dong et al., 2017; Liang et al., 2008; Lee et al., 2007). The main attributes assessed by the tasters include colour, aroma, taste and mouth feel of tea liquor and the appearance of dry tea (Zhu et al., 2017; Zheng et al., 2016b; Chen et al., 2015; Kumar et al., 2011a; Liang et al., 2003). Although sensory evaluation is quick and practical to use, it is limited since it requires identification and training to produce skilled and experienced professional tasters (Zhi et al., 2017; Stone et al., 2004; Liang et al., 2003) who are not easily found (Corley et al., 2005). It is also time consuming,

the tasters sometimes get exhausted and the approach is susceptible to many sources of variation because of individual tasters' preferences (Sinija et al., 2011; Bhattacharyya et al., 2008).

Chemical and physical analytical methods have also been developed for identifying biochemical components influencing tea quality (Yashin et al., 2015; Chen et al., 2015; Liang et al., 2008; Liang et al., 2003). Most of these techniques are objective, repeatable, reproducible and are not affected by fatigue (Zou et al., 2018; Yashin et al., 2015; Chen et al., 2015). However, these analytical techniques are expensive to acquire and maintain, require specialized expertise to operate and are not practical to use commercially. There is need for the tea industry to find a universally acceptable and objective method of evaluating tea quality (Zhu et al., 2017; Liang et al., 2003). With the current challenge of finding a reliable method of evaluating quality in tea, we review important biochemical compounds influencing quality in tea, sensory and objective methods that could be considered for evaluating the quality of made tea. This review will provide some insights on the current tea quality evaluation techniques and make suggestions for future research in the tea industry.

4.2. Tea quality attributes

4.2.1. Sensory attributes of tea

These are attributes that are described by tea tasters using the sensory organs such as tongue, nose and eye to perceive the taste, aroma and colour of tea, respectively. Flavour is comprised of taste, mouthfeel and aroma, while colour defines the appearance of dry tea and liquor (Ho et al., 2015). Taste, mouthfeel and colour are formed by non-volatile compounds while aroma is produced by volatile compounds in tea (Chen et al., 2008d; Kawakami, 1997; Yamanishi, 1995; Hara et al., 1995). Tea tasters use their experience and descriptions such as colour and shape of dried tea leaves, colour of infused tea, brisk, astringency, bitterness, mellowness and slight sweetness, thickness and fresh, fragrant, flowery, dull and grassy aroma to classify and grade manufactured teas (Lee et al., 2007).

4.3. Biochemical attributes and their influence on tea quality

Polyphenols consisting of flavonoids and phenolic acids are the main compounds known to influence quality in tea (Abdel-Rahman et al., 2011; Caffin et al., 2004; Harbowy et al., 1997; Balentine et al., 1997). Flavonoids are composed of flavan-3-ols (catechins), flavones, flavonols, flavanones, anthocyanins and isoflavonoids (Botten et al., 2015; Namal Senanayake, 2013; Kumar et al., 2011a; Hodgson et al., 2010; Mukhtar et al., 2000b). Flavan-3-ols (catechins) constitute between 60–80 % of the total polyphenols (Liang et al., 2003; Harbowy et al., 1997; Balentine et al., 1997) and they include (-)-epigallocatechin-3-gallate (EGCG), (-)-epigallocatechin (EGC), (-)-epicatechin-3- gallate (ECG) and epicatechin (EC) , (+)-gallocatechin (GC), (-)-gallocatechin gallate (GCG), (-)-catechin gallate (CG) and (+)-catechin (C) (El-Shahawi et al., 2012b; Chen et al., 2010; Friedman et al., 2005; Zeeb et al., 2000; Lee et al., 2000; Yamamoto et al., 1997).

Catechins influence tea quality by contributing to the characteristic properties of tea such as bitterness, astringency and sweet aftertaste (Scharbert et al., 2005; Scharbert et al., 2004; Mahindroo, 2000). The gallated catechins (EGCG and ECG) are astringent and give tea the bitter taste properties while the non-gallated catechins (C, EGC and EC) are less astringent and contribute to the sweet aftertaste of tea (Chen et al., 2014; Narukawa et al., 2011; Narukawa et al., 2010; Hayashi et al., 2010; Scharbert et al., 2005; Scharbert et al., 2004; Hara, 2001). Oolong teas are less astringent, but have a stronger sweet taste compared to green tea, while black tea has a combination of taste properties such as bitterness, astringency, sweetness, malty and green/grassy tastes (Alasalvar et al., 2012; Chaturvedula et al., 2011; Lee et al., 2007). Polyphenols are currently measured using chromatographic, spectroscopic and capillary electrophoresis techniques and can be used as quality indicators in tea.

4.4. Tea quality evaluation methods

4.4.1. Sensory evaluation

Sensory evaluation is a scientific method used to analyse the properties of food as perceived by the senses of sight, smell and taste (Olafsdottir et al., 1997). In tea quality determination, sensory evaluation relies on trained professional tasters who have developed their own language to describe various quality attributes of made tea (Zhu et al., 2017; Dong et al., 2017; Gill et al., 2011b; Bhuyan et al., 2009). The professional tasters use their experience and multiple sensory organs (eye, nose and tongue) to perceive the various aspects of tea quality (Fikri et al., 2011), such as appearance (colour and shape of dried tea leaves, colour of infused tea), taste (brisk, astringency, bitterness, mellowness and slight sweetness), mouthfeel (thickness) and aroma (fresh, fragrant, flowery, dull and grassy) of made tea (Yu et al., 2014; Buratti et al., 2013; Sinija et al., 2011; Bhattacharyya et al., 2008; Lee et al., 2007; Hara et al., 1995).

Sensory evaluation is the traditional method of evaluating tea quality commercially (Owuor et al., 2006; Stone et al., 2004; Liang et al., 2003), because it is easily acceptable, approximates consumers perceptions, it is quick and does not require the use of expensive equipment. However, it is time consuming, requires experienced trained tasters, expensive and the tasters get exhausted when analysing a large number of samples (Bhattacharyya et al., 2008; Yu et al., 2007; Okinda Owuor et al., 2006). It is also a highly subjective method (Bian et al., 2013; Okinda Owuor et al., 2006). Hence, an objective evaluation system is required for consistent evaluation of sensory tea quality (Zhu et al., 2017). Objective methods (chromatography, spectroscopy, capillary electrophoresis) measure the chemical compounds such as catechins, caffeine and theanine, that are known to correlate with sensory evaluation (Liang et al., 2003).

4.4.2. Spectroscopic methods

4.4.2.1. Mass spectrometry (MS)

Mass spectrometry is a powerful analytical technique used to quantify known materials, to identify unknown compounds within a sample, and to elucidate the structure and chemical properties of different

molecules (Griffiths, 2008). It measures the mass-to-charge ratios of ions when a sample is ionized (de Hoffmann, 2005). It consists of a vacuum system, sample introduction device, an ionization source, mass analyser and an ion detector (Steinmann et al., 2011). The complete process involves the conversion of the sample into gaseous ions, with or without fragmentation, which is then characterized by their mass to charge ratios (m/z) and relative abundances (de Hoffmann, 2005).

The first process of analysis using MS is ionization and it involves techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) (Awad et al., 2015; Domon et al., 2006). MS has been used in many studies to provide knowledge on the molecular mass and the structure of catechins and caffeine in tea (Poon, 1998). For instance, Chen et al. (2007a) used MS to successfully differentiate 40 tea samples including green tea, oolong tea, and jasmine tea by analysing the chemical fingerprints of the samples. Additionally, Menet et al. (2004) used MALDI-TOF mass Spectrometry to study the structure of theaflavins and thearubigins from tea extracts. SALDI-MS was also successfully applied using TiO2 NPs as selective probes and matrices to determine the concentrations of theanine, catechin, EGC, ECG, and EGCG in tea samples from Taiwan (Chen et al., 2013).

Liquid chromatography coupled with mass spectrometry (LC-MS) and Gas chromatography coupled with mass spectrometry (GC-MS) are the most preferred analytically techniques for tea volatiles and non-volatile analysis (Lee et al., 2000). LC-MS is a powerful and highly sensitive method that combines the physical separation abilities of liquid chromatography (LC) with the mass analytical capabilities of mass spectrometry (MS), and has been used to identify and quantify various polyphenols and amino acids in tea (Tao et al., 2016a; Sapozhnikova, 2014; Cordero et al., 2009; Wang et al., 2008a; Lin et al., 2008; Venzie et al., 2007; Kiehne et al., 1996). The main analysis conducted include identification of pesticides in tea samples (Huang et al., 2009), characterization of green tea (Clifford et al., 2007), classification of black tea (Sang et al., 2004), analysis of anthocyanin (Kiehne et al., 1997), analysis of Japanese tea quality (Pongsuwan et al., 2008a) and studying the influence of shade on flavonoid biosynthesis in tea (Wang et al., 2012).

GC-MS is the most preferable technique and has been used extensively for separating, identifying and quantifying volatile compounds in tea (Yang et al., 2018b; Ye et al., 2016; Xu et al., 2016; Lee et al., 2013; Pripdeevech et al., 2011; Wang et al., 2009). GC-MS has been used to study individual volatile compounds and in metabolic fingerprinting of green tea, oolong tea and black tea (Ye et al., 2012; Kumazawa et al., 2002; Kawakami et al., 1995). It has also been used in the characterisation of aroma from Pu-erh teas (Lv et al., 2012), discrimination of green teas from different geographical origins (Ye et al., 2012), analysis of agricultural residues on tea (Zhang et al., 2010), quality prediction of Japanese green tea (Pongsuwan et al., 2008a) and Longjing teas (Lin et al., 2012) and characterization of the chemical differences between solvent extracts from pu-erh tea and black tea (Gong et al., 2012).

Although MS, LC-MS and GC-MS are robust, accurate and highly sensitive techniques of analysing biochemical compounds in tea, they have not been used commercially in tea quality determination. They require skilled technical people to operate, are expensive, time consuming (require extra sample preparation techniques) and laborious.

4.4.2.2. Nuclear magnetic resonance (NMR)

NMR spectroscopy is a robust and non-destructive analytical technique that detects radiofrequency electromagnetic signals that are produced by the atomic nuclei within molecules to reveal information of samples (Bothwell et al., 2011). It has been used mainly for metabolite fingerprinting and measuring concentration of chemicals in complex compounds, such as polyphenols, amino acids and carbohydrates in plants (de B. Harrington et al., 2017; Larive et al., 2015). ^{1}H-NMR spectroscopy has been applied in chemical characterization and the simultaneous analysis of caffeine, gallic acid, theanine and catechins in commercial green teas (Yuan et al., 2014). It was also used to study the metabolic behaviour of tea during fermentation (Lee et al., 2011), finding the relationship between sensory evaluation of tea quality and metabolite quantities in green tea (Tarachiwin et al., 2007; Le Gall et al., 2004), quantitative determination of aluminium in tea (Koch, 1990), investigation of the metabolism of black tea via a nonspecific screening method (Daykin et al., 2005), studying the interaction of tea catechin and epigallocatechin gallate (EGCG),

with the model membrane of dimyristoylphosphatidylcholine (DMPC) (Kumazawa et al., 2004), quality assessment of green tea (Le Gall et al., 2004), discrimination by geographical origin of oolong tea (Meng et al., 2017), metabolite profiling of Japanese tea to evaluate their quality (Tarachiwin et al., 2007) and to investigate the effects of climatic conditions on green tea metabolites in three different growing areas of South Korea (Lee et al., 2010).

Although NMR spectroscopy is less sensitive compared to MS, it is more reproducible, non-destructive, requires minimal sample preparation steps, all metabolites can be detected in one measurement and it can also be used to identify unknown compounds in complex mixtures (Mahrous et al., 2015; Emwas, 2015). However, NMR has not been used practically in commercial tea quality evaluation because it is an expensive technique and requires skilled personnel.

4.4.2.3. Near infrared (NIR) spectroscopy

Near-infrared (NIR) spectroscopy is based on the absorption of electromagnetic radiation at wavelengths in the range 780–2500 nm (Chen et al., 2015). NIR spectroscopy is used routinely for the compositional, functional and sensory analysis of food ingredients, process intermediates and final products (Osborne, 2006). It has been applied in qualitative and quantitative research including the identification of geographical origins of teas (Zhuang et al., 2017; Ye et al., 2016; Chen et al., 2005), quantification of polyphenols, caffeine and amino acids (Zareef et al., 2018; Liu et al., 2018; Chen et al., 2018; Chanda et al., 2016; Lee et al., 2014b; Lin, 2009) and predicting the age of tea (Xu et al., 2011). Hyperspectral imaging (HSI) combines the qualities of an optical spectroscopy as an analytical tool with two-dimensional object visualization of optical imaging (Vasefi et al., 2016; Vo-Dinh, 2004). It produces a three-dimensional output containing spectral and spatial information, corresponding to the physical and chemical characteristics of a sample being analysed. Near infrared (NIR) hyperspectral imaging (HSI) has been used for the classification of six different commercial teas (oolong, green, yellow, white, black and Pu-erh) (Mishra et al., 2018b). NIR spectroscopy is a powerful analytical technique that can be used for routine quality evaluation because it is inexpensive, does not require sample preparation, is non-destructive, several

constituents can be measured simultaneously (Chen et al., 2018; Chen et al., 2015), it is rapid and does not use hazardous chemicals (Panigrahi et al., 2016; Lee et al., 2014b). However, the main constraints of NIR spectroscopy is its reliance on reference methods and model development using chemometrics, which requires highly skilled expertise (Manley, 2014).

4.4.3. Chromatographic methods

4.4.3.1. High performance liquid chromatography (HPLC)

HPLC is an analytical separation technique that utilize a liquid mobile phase and a solid stationary phase to isolate and measure the quantities of compounds in samples (Forgács et al., 2003). The samples are separated based on their interaction with the solid and mobile phases. HPLC consists of a pump, injector, chromatographic column, thermostatically controlled oven, detector and computer for data analysis (Yang et al., 2007; Forgács et al., 2003). The main detectors in HPLC include ultraviolet-visible (UV-vis), fluorescence, refractive index, electrochemical and mass spectrometry detectors (Kumamoto et al., 2000). HPLC is the most common method of evaluating polyphenols, alkaloids, amino acids and carbohydrates in tea (Yashin et al., 2015; Novak et al., 2010; Lv et al., 2009; Cordero et al., 2009; Yamauchi et al., 2008; Pelillo et al., 2004; Sano et al., 2001; Wu et al., 1998). Reversed phase HPLC with UV-absorbance detection is the most common and frequently used technique for estimating tea catechins and caffeine (Yashin et al., 2015).

In tea analysis, HPLC has been exploited for analysing of polyphenols in green and black tea (Del Rio et al., 2004), determination of catechins, caffeine, gallic acids (Wang et al., 2000a; Khokhar et al., 1997), identification of green tea's quality by measuring catechins and caffeine contents (Chen et al., 2008a), quality assessment and quantitative analysis of flavonoids from tea samples of different origins (Sultana et al., 2008), comparison of catechins and purine alkaloids in albino and normal green tea cultivars (Wei et al., 2012), chemical fingerprinting for quality control and identification of green tea (He et al.,

2015) and estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions (Liang et al., 2003).

HPLC is a very accurate and reliable method, commonly used for quantifying non-volatile compounds in tea. However, it is expensive to acquire and maintain, requires technical expertise to operate and the steps involved in HPLC analysis are laborious and time consuming (Kim et al., 2007).

4.4.3.2. Gas Chromatography (GC)

Gas chromatography (GC) is a separation technique using gas flow through a glass or metal column that separates compounds based on both volatility and interaction with the liquid stationary phase. GC is frequently used to measure volatile compounds in tea, and consists of a mobile phase (gas) and the stationary phase (solid or liquid) (Forgács et al., 2003; Niessen, 2001). Gas chromatography can either be GSC (gas–solid chromatography) or gas–liquid chromatography (GLC). In GSC, the stationary phase consists of a solid material such as silica, alumina or carbon, and is commonly used for separating permanent gases and low-boiling hydrocarbons. For GLC, the stationary phase is a liquid, while the mobile phase is a gas (Seneca, 2007). The main component of GC includes; a carrier gas system, injector, gas chromatographic column, detector and a computer for processing data. GC is the main method of evaluating volatile compounds in tea (Schuh et al., 2006; Xu et al., 2002; Guth et al., 1993; Owuor, 1992; Owuor et al., 1988). Other modifications of GC are also available and have been used for quantifying flavour compounds in tea and they include; headspace gas chromatography (Headspace GC) (Baptista et al., 1998), GC coupled to solid phase microextraction (SPME-GC) (Wang et al., 2008b; Reto et al., 2007), GC with electron capture detector (GC-ECD) (Gu et al., 2011), GC with flame ionization detector (GC-FID) (Reto et al., 2007).

In tea, GC has been used for aroma characterisation of Pu-erh tea (Lv et al., 2012), ranking of Japanese green tea (Pongsuwan et al., 2008b), characterization of volatile components of orthodox black tea (Rawat et al., 2007), pattern recognition of green, black and oolong teas (Togari et al., 1995), separating and identifying individual catechins (Dalluge et al., 2000) and comparison of teas (Shellie et al., 2000).

Although GC is an accurate and reliable method of quantifying volatile compounds in tea, it is an expensive method, requires skilled personnel and has many time-consuming steps.

4.4.3.3. Capillary electrophoresis (CE)

Capillary electrophoresis separates ions when an electric charge is applied to a sample in a capillary tube. It separates ions according to their mass to charge ratio, as they move in the capillary tube. The main methods used in CE are capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), micellar electrokinetic capillary chromatography (MEKC), capillary isoelectring focusing and capillary isotachophoresis (Forgács et al., 2003; Wright, 2002). CE and HPLC are the main techniques used for separating polyphenols, alkaloids and amino acids in tea (Ning et al., 2010; Li et al., 2009; Lee et al., 2000).

CE has been used for tea quality estimation (Horie et al., 1998), comparative analysis of catechins and theaflavins (Lee et al., 2000), determination of anti-carcinogenic polyphenols in green tea (Arce et al., 1998), determination of amino acids in tea leaves (Hsieh et al., 2007), analysis of black tea theaflavins (Wright et al., 2001) and analysis of organic anions in tea infusions (Horie et al., 1998). CE is a reliable and reproducible method of evaluating non-volatile compounds in tea. However, CE is expensive and requires skilled personnel (Li et al., 2009).

4.4.4. Sensory bionics techniques

Instruments that use sensory bionics technologies have sensory arrays that are sensitive to different chemicals and have advanced pattern recognition system for reliable and automatic signal processing. These instruments include electronic nose, electronic tongue, computer vision and multiple sensor systems which have been used to determine the quality of tea based on the sense of smell, taste, vision and a combination of senses, respectively (Zhi et al., 2017; Chen et al., 2015).

4.4.4.1. Electronic nose (E-Nose)

An electronic nose is an instrument that comprises an array of chemical sensors and data processing components for data recording and analysis (Arshak et al., 2004). It has a pattern recognition system that mimics the sense of smell in mammalian biological systems. It detects, discriminates and classifies different complex aromas (Chen et al., 2015; Röck et al., 2008; Bhattacharyya et al., 2005; Gardner et al., 1994). It is made up of a pump, inlet sampling system, array of sensors and signal processing system (Pavlou et al., 2004). E-noses analyses aroma profiles by registering signals produced by different volatile compounds and comparing the pattern of responses produced by the samples being analysed (Rodríguez-Méndez et al., 2016; Strike et al., 1999; Gardner et al., 1994).

The main sensors used in electronic noses include metal oxide (MOX) (Skov et al., 2005; James et al., 2005), metal oxide semiconductor field effect transistors (MOSFET) (Smyth et al., 2012), conducting polymer (CP), (Zoecklein et al., 2011), optical (Elosua et al., 2012), quartz microbalance (QMB) (Zampetti et al., 2008) and surface acoustic wave sensors (SAW) (García et al., 2006). In tea, an electronic noses are used for evaluating the quality of different teas (Tozlu et al., 2018; Zhi et al., 2017; Buratti et al., 2013; Yu et al., 2009; Tudu et al., 2009; Yu et al., 2008), categorization of aroma and flavour compounds in black tea (Kawakami et al., 2004), classification of tea (Roy et al., 2013; Roy et al., 2012; Yu et al., 2008; Yu et al., 2007), determination of storage time (Yu et al., 2009) and monitoring of black tea fermentation process (Bhattacharyya et al., 2007). All these applications confirm the benefits of adopting an E-nose in routine tea quality evaluation which include; reliability, ease of operation, cost effectiveness, rapidity and accuracy (Chen et al., 2015; Chen et al., 2011a). Therefore, an E-nose could complement or even replace the expensive and time-consuming gas chromatography.

4.4.4.2. Electronic tongue (Artificial tongue)

An electronic tongue (E-tongue) is an analytical instrument which artificially reproduces the taste sensation (Escuder-Gilabert et al., 2010). It consists of low selective sensor arrays which are capable of distinguishing various samples, a pattern recognition and multivariate calibration systems for data

processing (Sliwinska et al., 2014; Tahara et al., 2013). The main sensors used in an electronic tongue system are voltammetric (Ciosek et al., 2007), amperometric (Buratti et al., 2013), impedentiometric (Cortina-Puig et al., 2007), acoustic wave (Sehra et al., 2004), optical (Sohn et al., 2005) and ion selective field effect transistors (ISFET) sensors (Moreno et al., 2006). An electronic tongue has been used for the determination of taste and mouth feel of black tea (Palit et al., 2010; Ivarsson et al., 2001) and green tea (Wu et al., 2011; Chen et al., 2011a; Lvova et al., 2003), classification of different teas (Wu et al., 2006; Scampicchio et al., 2006) and tea grades (Zhi et al., 2014; Xiao et al., 2009; Chen et al., 2008d). Several studies have shown correlations between E-tongues and sensory evaluation results (Scampicchio et al., 2006; Parra et al., 2006), hence it has great potential to be adopted in practical commercial tea quality evaluation because it is cost effective, reliable and simple to use (Zhi et al., 2017).

4.4.4.3. Computer vision

Computer vision is a technique that analyses the image of samples accurately using computers based on the physical properties such as colour and texture (Gill et al., 2011a). The colour of made tea is measured and accurately characterized by capturing sample images with a machine vision acquisition system and extracting the colour features with digitization (Dana et al., 2008). Classification of samples is then conducted using statistical and modelling approaches (Gill et al., 2011a). Compared to other techniques, computer vision systems are cost effective, consistent, fast and accurate (Chen et al., 2015; Gill et al., 2011a). Computer vision has been applied in the monitoring and grading of tea (Gill et al., 2011a), identification of tea varieties (Chen et al., 2008b), tea quality evaluation (Dong et al., 2017; Sharma et al., 2013; Rakhmawati et al.) and estimation of theaflavins and thearubigins in tea samples (Akuli et al., 2016). Computer vision is efficient in quality evaluation (Anami et al., 2009; Alfatni et al., 2008), because it is automated, non-destructive and cost effective (Sun, 2016), and therefore could be applied for routine quality evaluation by the tea industry because it is more objective and reliable.

4.4.4.4. Multiple sensor system

This involves fusion of several sensors such as E-nose, E-tongue and computer vision. It mimics the human sensory system that use multiple mammal sensory organs (eye, nose, and tongue) to perceive the various sensory properties such as colour and flavour, and gives a final score which comprehensively define the quality of tea (Kiani et al., 2016; Fikri et al., 2011; Huang et al., 2004). A system with multiple sensors could improve the accuracy of determining quality properties compared to an individual system (Li et al., 2018; Chen et al., 2015; Hong et al., 2014; Haddi et al., 2014; Cole et al., 2011). Fikri et al. (2011) developed a human sensory imitating system that combines an E-nose and E-tongue and accurately identified four different types of teas. Similarly, Roy et al. (2012) accurately evaluated the quality of tea using a system that combined E-nose and E-tongue. Consequently, multiple sensor system could be considered for commercial quality evaluation to determine the colour, taste, mouth and aroma of made tea.

4.5. Future trends of tea quality assessment

Tea quality is evaluated traditionally by professional tea tasters and the results are highly subjective (Bian et al., 2012). Analytical approaches such as chromatography, spectroscopy and capillary electrophoresis have been developed and are commonly used to determine polyphenols, caffeine and amino acids, which influence tea quality (Koch et al., 2018; Punyasiri et al., 2017; Owuor et al., 2007). However, most of these methods are time-consuming, tedious, destructive, require highly skilled personnel and expensive (Xiong et al., 2015; Ren et al., 2013).

Sensory bionics techniques (E-nose, E-tongue, computer vision and multiple sensor) have recently been developed and can mimic the sensory perception of human smell, taste and sight, and have been used to determine the quality of tea (Zhi et al., 2017; Chen et al., 2015). Compared to sensory evaluation and the physical and chemical methods (chromatography, spectroscopy, capillary electrophoresis), the sensory bionics techniques are non-destructive, cost effective, easy to use and rapid methods, and could be applied practically in routine tea quality evaluation. Additionally, these techniques are highly effective as they do not require sample preparation and the use of dangerous chemicals.

4.6. Conclusion

Recent advances in bionics, and physical and chemical methods provide opportunities for adopting a more reliable and objective method of tea quality evaluation. Evaluation of tea quality is an important process for tea companies and tea buyers. The traditional sensory evaluation method using professional tasters is still dominant in commercial tea quality evaluation. However, sensory evaluation is a highly subjective, unreliable and time-consuming method. Although spectrophotometric, chromatographic and capillary electrophoresis techniques are routinely used to quantify catechins, caffeine and amino acids for tea quality determination, they are expensive, require skilled personnel and are not practical for use commercially. Recently, bionics have been developed and they imitate human senses of taste, smell, sight and texture, and have the potential to complement or replace the current sensory evaluation methods. These devices are cost effective, easy to use, reliable, accurate and their output correlate with human perception of colour and flavor. However, more research and effort are required to facilitate the early adoption and application of these new techniques in commercial tea quality evaluation.

References

Abdel-Rahman, A., Anyangwe, N., Carlacci, L., Casper, S., Danam, R. P., Enongene, E., Erives, G., Fabricant, D., Gudi, R., Hilmas, C. J. (2011). The safety and regulation of natural products used as foods and food ingredients. *Toxicological Sciences, 123*(2), 333-348.

Akuli, A., Pal, A., Bej, G., Dey, T., Ghosh, A., Tudu, B., Bhattacharyya, N., Bandyopadhyay, R. (2016). A MACHINE VISION SYSTEM FOR ESTIMATION OF THEAFLAVINS AND THEARUBIGINS IN ORTHODOX BLACK TEA. *International Journal on Smart Sensing & Intelligent Systems, 9*(2).

Alasalvar, C., Topal, B., Serpen, A., Bahar, B., Pelvan, E., Gökmen, V. (2012). Flavor characteristics of seven grades of black tea produced in Turkey. *Journal of Agricultural and Food Chemistry, 60*(25), 6323-6332.

Alfatni, M. S. M., Shariff, A. R. M., Shafri, H. Z. M., Saaed, O. M. B., Eshanta, O. M. (2008). Oil palm fruit bunch grading system using red, green and blue digital number. *Journal of Applied Sciences, 8*(8), 1444-1452.

Anami, B. S., Burkpalli, V. C. (2009). Color based identification and classification of boiled food grain images. *International journal of food engineering, 5*(5).

Arshak, K., Moore, E., Lyons, G. M., Harris, J., Clifford, S. (2004). A review of gas sensors employed in electronic nose applications. *Sensor review, 24*(2), 181-198.

Awad, H., Khamis, M. M., El-Aneed, A. (2015). Mass spectrometry, review of the basics: ionization. *Applied Spectroscopy Reviews, 50*(2), 158-175.

Balentine, D. A., Wiseman, S. A., Bouwens, L. C. M. (1997). The chemistry of tea flavonoids. *Critical Reviews in Food Science & Nutrition, 37*(8), 693-704.

Baptista, J. A. B., da P Tavares, J. F., Carvalho, R. C. B. (1998). Comparison of catechins and aromas among different green teas using HPLC/SPME-GC. *Food Research International, 31*(10), 729-736.

Bhattacharyya, N., Bandyopadhyay, R., Bhuyan, M., Tudu, B., Ghosh, D., Jana, A. (2008). Electronic nose for black tea classification and correlation of measurements with "Tea Taster" marks. *IEEE transactions on instrumentation and measurement, 57*(7), 1313-1321.

Bhattacharyya, N., Bandyopadhyay, R., Bhuyan, M., Ghosh, A., Mudi, R. K. (2005). Correlation of multi-sensor array data with 'tasters' panel evaluation for objective assessment of black tea flavour. *Int. Proc. ISOEN-2005, Barcelona, Spain*, 13-15.

Bhattacharyya, N., Seth, S., Tudu, B., Tamuly, P., Jana, A., Ghosh, D., Bandyopadhyay, R., Bhuyan, M. (2007). Monitoring of black tea fermentation process using electronic nose. *Journal of Food Engineering, 80*(4), 1146-1156. doi:10.1016/j.jfoodeng.2006.09.006

Bhuyan, L. P., Hussain, A., Tamuly, P., Gogoi, R. C., Bordoloi, P. K., Hazarika, M. (2009). Chemical characterisation of CTC black tea of northeast India: correlation of quality parameters with tea tasters' evaluation. *Journal of the Science of Food and Agriculture, 89*(9), 1498-1507. doi:10.1002/jsfa.3614

Bian, M., Skidmore, A. K., Schlerf, M., Liu, Y., Wang, T. (2012). Estimating biochemical parameters of tea (Camellia sinensis (L.)) using hyperspectral technique. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39*, B8.

Bian, M., Skidmore, A. K., Schlerf, M., Wang, T., Liu, Y., Zeng, R., Fei, T. (2013). Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels. *ISPRS journal of photogrammetry and remote sensing, 78*, 148-156.

Bothwell, J. H. F., Griffin, J. L. (2011). An introduction to biological nuclear magnetic resonance spectroscopy. *Biological Reviews, 86*(2), 493-510.

Botten, D., Fugallo, G., Fraternali, F., Molteni, C. (2015). Structural Properties of Green Tea Catechins. *J Phys Chem B, 119*(40), 12860-12867. doi:10.1021/acs.jpcb.5b08737

Buratti, S., Casiraghi, A., Minghetti, P., Giovanelli, G. (2013). The joint use of electronic nose and electronic tongue for the evaluation of the sensorial properties of green and black tea infusions as related to their chemical composition. *Food and Nutrition Sciences, 4*(6), 605-615.

Caffin, N., D'Arcy, B., Yao, L., Rintoul, G. 2004. Developing an index of quality for Australian tea, Rural Industries Research and Development Corporation. In.

Chanda, S., De, A., Tudu, B., Bandyopadhyay, R., Hazarika, A. K., Sabhapondit, S., Baruah, B. D., Tamuly, P., Bhattachryya, N. (2016, 21-23 Oct. 2016). *Prediction of polyphenol content in tea leaves using NIR spectroscopy.* Paper presented at the 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI).

Chaturvedula, V. S. P., Prakash, I. (2011). The aroma, taste, color and bioactive constituents of tea. *Journal of Medicinal Plants Research, 5*(11), 2110-2124.

Chen, G. H., Yang, C. Y., Lee, S. J., Wu, C. C., Tzen, J. T. C. (2014). Catechin content and the degree of its galloylation in oolong tea are inversely correlated with cultivation altitude. *J Food Drug Anal, 22*(3), 303-309. doi:10.1016/j.jfda.2013.12.001

Chen, Q., Zhao, J., Cai, J. (2008). Identification of tea varieties using computer vision. *Transactions of the ASABE, 51*(2), 623-628.

Chen, Q., Zhao, J., Guo, Z., Wang, X. (2010). Determination of caffeine content and main catechins contents in green tea (Camellia sinensis L.) using taste sensor technique and multivariate calibration. *Journal of Food Composition and Analysis, 23*(4), 353-358.

Chen, Q., Zhao, J., Vittayapadung, S. (2008). Identification of the green tea grade level using electronic tongue and pattern recognition. *Food Research International, 41*(5), 500-504. doi:10.1016/j.foodres.2008.03.005

Chen, Q., Zhao, J., Chen, Z., Lin, H., Zhao, D.-A. (2011). Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. *Sensors and actuators B: Chemical, 159*(1), 294-300.

Chen, Q., Chen, M., Liu, Y., Wu, J., Wang, X., Ouyang, Q., Chen, X. (2018). Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea. *Journal of food science and technology, 55*(10), 4363-4368. doi:10.1007/s13197-018-3353-1

Chen, Q., Zhang, D., Pan, W., Ouyang, Q., Li, H., Urmila, K., Zhao, J. (2015). Recent developments of green analytical techniques in analysis of tea's quality and nutrition. *Trends in Food Science & Technology, 43*(1), 63-82. doi:10.1016/j.tifs.2015.01.009

Chen, Q., Zhao, J., Zhang, H., Muhua, L., Fang, M. (2005). Qualitative Identification of Tea by near Infrared Spectroscopy Based on Soft Independent Modelling of Class Analogy Pattern Recognition. *Journal of Near Infrared Spectroscopy, 13*(6), 327-332.

Ciosek, P., Wroblewski, W. (2007). Sensor arrays for liquid sensing–electronic tongue systems. *Analyst, 132*(10), 963-978.

Cole, M., Covington, J. A., Gardner, J. W. (2011). Combined electronic nose and tongue for a flavour sensing system. *Sensors and actuators B: Chemical, 156*(2), 832-839.

Cordero, C., Canale, F., Rio, D. D., Bicchi, C. (2009). Identification, quantitation, and method validation for flavan-3-ols in fermented ready-to-drink teas from the Italian market using HPLC-UV/DAD and LC-MS/MS. *Journal of Separation Science, 32*(21), 3643-3651.

Corley, R. H. V., Chomboi, K. C. (2005). Tea tasting-a statistical evaluation of tasters' skills. *Tea, 26*(1), 10-18.

Cortina-Puig, M., Muñoz-Berbel, X., Alonso-Lomillo, M. A., Muñoz-Pascual, F. J., del Valle, M. (2007). EIS multianalyte sensing with an automated SIA system—An electronic tongue employing the impedimetric signal. *Talanta, 72*(2), 774-779.

Dana, W., Ivo, W. (2008). Computer image analysis of seed shape and seed color for flax cultivar description. *Computers and electronics in agriculture, 61*(2), 126-135.

de B. Harrington, P., Wang, X. (2017). Spectral Representation of Proton NMR Spectroscopy for the Pattern Recognition of Complex Materials. *Journal of Analysis and Testing, 1*(2), 10. doi:10.1007/s41664-017-0003-y

de Hoffmann, E. (2005). Mass Spectrometry. In *Kirk-Othmer Encyclopedia of Chemical Technology*.

Domon, B., Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. *Science, 312*(5771), 212-217. doi:10.1126/science.1124619

Dong, C.-w., Zhu, H.-k., Zhao, J.-w., Jiang, Y.-w., Yuan, H.-b., Chen, Q.-s. (2017). Sensory quality evaluation for appearance of needle-shaped green tea based on computer vision and nonlinear tools 基于机器视觉和非线性的芽形绿茶外形感官品质评价. *Journal of Zhejiang University-SCIENCE B, 18*(6), 544-548.

El-Shahawi, M. S., Hamza, A., Bahaffi, S. O., Al-Sibaai, A. A., Abduljabbar, T. N. (2012). Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography. *Food Chem, 134*(4), 2268-2275. doi:10.1016/j.foodchem.2012.03.039

Elosua, C., Bariain, C., Luquin, A., Laguna, M., Matias, I. R. (2012). Optical fiber sensors array to identify beverages by their odor. *IEEE Sensors Journal, 12*(11), 3156-3162.

Emwas, A.-H. M. (2015). The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In J. T. Bjerrum (Ed.), *Metabonomics: Methods and Protocols* (pp. 161-193). New York, NY: Springer New York.

Escuder-Gilabert, L., Peris, M. (2010). Highlights in recent applications of electronic tongues in food analysis. *Analytica Chimica Acta, 665*(1), 15-25.

Fikri, N. A., Adom, A. H., Ahmad, M. N., Abdullah, A. H., Zakaria, A., Markom, M. A. (2011). Development of human sensory mimicking system. *Sensor Letters, 9*(1), 423-427.

Forgács, E., Cserháti, T. (2003). CHROMATOGRAPHY | Principles. In B. Caballero (Ed.), *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (pp. 1259-1267). Oxford: Academic Press.

Friedman, M., Kim, S. Y., Lee, S. J., Han, G. P., Han, J. S., Lee, K. R., Kozukue, N. (2005). Distribution of catechins, theaflavins, caffeine, and theobromine in 77 teas consumed in the United States. *Journal of food science, 70*(9), C550-C559.

Gallaher, R. N., Gallaher, K., Marshall, A. J., Marshall, A. C. (2006). Mineral analysis of ten types of commercially available tea. *Journal of Food Composition and Analysis, 19*, S53-S57.

García, M., Fernández, M. J., Fontecha, J. L., Lozano, J., Santos, J. P., Aleixandre, M., Sayago, I., Gutiérrez, J., Horrillo, M. C. (2006). Differentiation of red wines using an electronic nose based on surface acoustic wave devices. *Talanta, 68*(4), 1162-1165.

Gardner, J. W., Bartlett, P. N. (1994). A brief history of electronic noses. *Sensors and actuators B: Chemical, 18*(1-3), 210-211.

Ghosh, A., Tamuly, P., Bhattacharyya, N., Tudu, B., Gogoi, N., Bandyopadhyay, R. (2012). Estimation of theaflavin content in black tea using electronic tongue. *Journal of Food Engineering, 110*(1), 71-79.

Gill, G. S., Kumar, A., Agarwal, R. (2011a). Monitoring and grading of tea by computer vision – A review. *Journal of Food Engineering, 106*(1), 13-19. doi:https://doi.org/10.1016/j.jfoodeng.2011.04.013

Gill, G. S., Kumar, A., Agarwal, R. (2011b). Monitoring and grading of tea by computer vision–A review. *Journal of Food Engineering, 106*(1), 13-19.

Griffiths, J. (2008). A brief history of mass spectrometry. *Anal Chem, 80*(15), 5678-5683.

Gu, X., Yao, C., Zhang, Z., Wan, X., Ning, J., Shao, W. (2011). GC–ECD method for determination of glucosidically bound aroma precursors in fresh tea leaves. *Chromatographia, 73*(1-2), 189-193.

Guth, H., Grosch, W. (1993). Identification of potent odourants in static headspace samples of green and black tea powders on the basis of aroma extract dilution analysis (AEDA). *Flavour and fragrance journal, 8*(4), 173-178.

Haddi, Z., Mabrouk, S., Bougrini, M., Tahri, K., Sghaier, K., Barhoumi, H., El Bari, N., Maaref, A., Jaffrezic-Renault, N., Bouchikhi, B. (2014). E-Nose and e-Tongue combination for improved recognition of fruit juice samples. *Food Chemistry, 150*, 246-253.

Hara, Y. (2001). *Green tea: health benefits and applications*: CRC press.

Hara, Y., Luo, S. J., Wickremashinghe, R. L., Yamanishi, T. (1995). IV. Biochemistry of processing black tea. *Food Reviews International, 11*, 457-471.

Harbowy, M. E., Balentine, D. A., Davies, A. P., Cai, Y. (1997). Tea chemistry. *Critical reviews in plant sciences, 16*(5), 415-480.

Hayashi, N., Chen, R., Hiraoka, M., Ujihara, T., Ikezaki, H. (2010). β-Cyclodextrin/surface plasmon resonance detection system for sensing bitter-astringent taste intensity of green tea catechins. *Journal of Agricultural and Food Chemistry, 58*(14), 8351-8356.

Ho, C.-T., Zheng, X., Li, S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9-27.

Hodgson, J. M., Croft, K. D. (2010). Tea flavonoids and cardiovascular health. *Molecular aspects of medicine, 31*(6), 495-502.

Hong, X., Wang, J. (2014). Detection of adulteration in cherry tomato juices based on electronic nose and tongue: Comparison of different data fusion approaches. *Journal of Food Engineering, 126*, 89-97.

Huang, Y.-b., Lan, Y.-b., Lacey, R. E. (2004). Artificial senses for characterization of food quality. *Journal of Bionic Engineering, 1*(4), 159-173. doi:10.1007/bf03399472

Ivarsson, P., Holmin, S., Höjer, N.-E., Krantz-Rülcker, C., Winquist, F. (2001). Discrimination of tea by means of a voltammetric electronic tongue and different applied waveforms. *Sensors and actuators B: Chemical, 76*(1-3), 449-454.

James, D., Scott, S. M., Ali, Z., O'hare, W. T. (2005). Chemical sensors for electronic nose systems. *Microchimica Acta, 149*(1-2), 1-17.

John, K. M. M., Thiruvengadam, M., Enkhtaivan, G., Kim, D. H. (2014). Variation in major phenolic compounds and quality potential of CTC black tea elicited by Saccharomyces cercevisiae and its correlation with antioxidant potential. *Industrial Crops and Products, 55*, 289-294.

Kawakami, M., Ganguly, S. N., Banerjee, J., Kobayashi, A. (1995). Aroma composition of oolong tea and black tea by brewed extraction method and characterizing compounds of Darjeeling tea aroma. *Journal of Agricultural and Food Chemistry, 43*(1), 200-207.

Kawakami, M., Sarma, S., Himizu, K., Konishi, Y., Kobayashi, A. (2004, 2004). *Aroma characteristics of Darjeeling tea*.

Kawakami, M. (1997). Comparison of extraction techniques for characterizing tea aroma and analysis of tea by GC-FTIR-MS. In *Plant Volatile Analysis* (pp. 211-229): Springer.

Kiani, S., Minaei, S., Ghasemi-Varnamkhasti, M. (2016). Fusion of artificial senses as a robust approach to food quality assessment. *Journal of Food Engineering, 171*, 230-239.

Kiehne, A., Engelhardt, U. H. (1996). Thermospray-LC-MS analysis of various groups of polyphenols in tea. *Zeitschrift für Lebensmittel-Untersuchung und Forschung, 202*(1), 48-54.

Kim, K. S., Park, S. H., Choung, M. G. (2007). Nondestructive determination of oil content and fatty acid composition in perilla seeds by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 55*(5), 1679-1685.

Koch, W., Kukula-Koch, W., Komsta, L., Marzec, Z., Szwerc, W., Glowniak, K. (2018). Green Tea Quality Evaluation Based on Its Catechins and Metals Composition in Combination with Chemometric Analysis. *Molecules, 23*(7). doi:10.3390/molecules23071689

Kumamoto, M., Sonda, T., Takedomi, K., Tabata, M. (2000). Enhanced separation and elution of catechins in HPLC using mixed-solvents of water, acetonitrile and ethyl acetate as the mobile phase. *Analytical Sciences, 16*(2), 139-144.

Kumar, P. V., Basheer, S., Ravi, R., Thakur, M. S. (2011). Comparative assessment of tea quality by various analytical and sensory methods with emphasis on tea polyphenols. *J Food Sci Technol, 48*(4), 440-446. doi:10.1007/s13197-010-0178-y

Kumar, R. S. S., Murugesan, S., Kottur, G., Gyamfi, D. (2013). Black tea: The plants, processing/manufacturing and production. In *Tea in health and disease prevention* (pp. 41-57): Elsevier.

Kumazawa, K., Masuda, H. (2002). Identification of potent odorants in different green tea varieties using flavor dilution technique. *Journal of Agricultural and Food Chemistry, 50*(20), 5660-5663.

Larive, C. K., Barding, G. A., Dinges, M. M. (2015). NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Analytical Chemistry, 87*(1), 133-146. doi:10.1021/ac504075g

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

Lee, B.-L., Ong, C.-N. (2000). Comparative analysis of tea catechins and theaflavins by high-performance liquid chromatography and capillary electrophoresis. *Journal of Chromatography A, 881*(1-2), 439-447. doi:10.1016/s0021-9673(00)00215-6

Lee, J.-E., Lee, B.-J., Chung, J.-O., Shin, H.-J., Lee, S.-J., Lee, C.-H., Hong, Y.-S. (2011). 1H NMR-based metabolomic characterization during green tea (Camellia sinensis) fermentation. *Food Research International, 44*(2), 597-604. doi:https://doi.org/10.1016/j.foodres.2010.12.004

Lee, J., Chambers, D. H. (2007). A lexicon for flavor descriptive analysis of green tea. *Journal of Sensory Studies, 22*(3), 256-272.

Lee, J., Chambers, D., Chambers, E., Adhikari, K., Yoon, Y. (2013). Volatile Aroma Compounds in Various Brewed Green Teas. *Molecules, 18*(8), 10024-10041. doi:10.3390/molecules180810024

Lee, M. S., Hwang, Y. S., Lee, J., Choung, M. G. (2014). The characterization of caffeine and nine individual catechins in the leaves of green tea (Camellia sinensis L.) by near-infrared reflectance spectroscopy. *Food Chem, 158*, 351-357. doi:10.1016/j.foodchem.2014.02.127

Li, L., Xie, S., Ning, J., Chen, Q., Zhang, Z. (2018). Evaluating green tea quality based on multisensor data fusion combining hyperspectral imaging and olfactory visualization systems. *Journal of the Science of Food and Agriculture, 0*(0). doi:doi:10.1002/jsfa.9371

Li, M., Zhou, J., Gu, X., Wang, Y., Huang, X., Yan, C. (2009). Quantitative capillary electrophoresis and its application in analysis of alkaloids in tea, coffee, coca cola, and theophylline tablets. *Journal of Separation Science, 32*(2), 267-274.

Li, S., Lo, C. Y., Pan, M. H., Lai, C. S., Ho, C. T. (2013). Black tea: chemical analysis and stability. *Food Funct, 4*(1), 10-18. doi:10.1039/c2fo30093a

Liang, Y., Lu, J., Zhang, L., Wu, S., Wu, Y. (2003). Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions. *Food Chemistry, 80*(2), 283-290. doi:https://doi.org/10.1016/S0308-8146(02)00415-6

Liang, Y. R., Ye, Q., Jin, J., Liang, H., Lu, J. L., Du, Y. Y., Dong, J. J. (2008). Chemical and instrumental assessment of green tea sensory preference. *International Journal of Food Properties, 11*(2), 258-272.

Lin, L.-Z., Chen, P., Harnly, J. M. (2008). New phenolic components and chromatographic profiles of green and fermented teas. *Journal of Agricultural and Food Chemistry, 56*(17), 8130-8140.

Lin, X. (2009). Qualitative and quantitative analysis method of tea by near infrared spectroscopy. *Spectroscopy and Spectral Analysis, 29*(9), 2417-2420.

Liu, Z., Xie, H.-l., Chen, L., Huang, J.-h. (2018). An Improved Weighted Partial Least Squares Method Coupled with Near Infrared Spectroscopy for Rapid Determination of Multiple Components and Anti-Oxidant Activity of Pu-Erh Tea. *Molecules, 23*(5), 1058.

Lv, Y., Yang, X., Zhao, Y., Ruan, Y., Yang, Y., Wang, Z. (2009). Separation and quantification of component monosaccharides of the tea polysaccharides from Gynostemma pentaphyllum by HPLC with indirect UV detection. *Food Chemistry, 112*(3), 742-746.

Lvova, L., Legin, A., Vlasov, Y., Cha, G. S., Nam, H. (2003). Multicomponent analysis of Korean green tea by means of disposable all-solid-state potentiometric electronic tongue microsystem. *Sensors and actuators B: Chemical, 95*(1-3), 391-399.

Magagna, F., Cordero, C., Cagliero, C., Liberto, E., Rubiolo, P., Sgorbini, B., Bicchi, C. (2017). Black tea volatiles fingerprinting by comprehensive two-dimensional gas chromatography - Mass spectrometry combined with high concentration capacity sample preparation techniques: Toward a fully automated sensomic assessment. *Food Chem, 225*, 276-287. doi:10.1016/j.foodchem.2017.01.003

Mahindroo, N. (2000). Tea-Camellia sinensis-report. *ICS-UNIDO, Trieste, Italy*.

Mahrous, E. A., Farag, M. A. (2015). Two dimensional NMR spectroscopic approaches for exploring plant metabolome: a review. *Journal of advanced research, 6*(1), 3-15.

Mishra, P., Nordon, A., Tschannerl, J., Lian, G., Redfern, S., Marshall, S. (2018). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering*.

Moreno, L., Merlos, A., Abramova, N., Jimenez, C., Bratov, A. (2006). Multi-sensor array used as an "electronic tongue" for mineral water analysis. *Sensors and actuators B: Chemical, 116*(1-2), 130-134.

Mukhtar, H., Ahmad, N. (2000). Tea polyphenols: prevention of cancer and optimizing health–. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.

Namal Senanayake, S. P. J. (2013). Green tea extract: Chemistry, antioxidant properties and food applications – A review. *Journal of Functional Foods, 5*(4), 1529-1541. doi:10.1016/j.jff.2013.08.011

Narukawa, M., Kimata, H., Noga, C., Watanabe, T. (2010). Taste characterisation of green tea catechins. *International Journal of Food Science & Technology, 45*(8), 1579-1585. doi:10.1111/j.1365-2621.2010.02304.x

Narukawa, M., Noga, C., Ueno, Y., Sato, T., Misaka, T., Watanabe, T. (2011). Evaluation of the bitterness of green tea catechins by a cell-based assay with the human bitter taste receptor hTAS2R39. *Biochemical and biophysical research communications, 405*(4), 620-625.

Niessen, W. M. A. (2001). Principles and instrumentation of gas chromatography-mass spectrometry. In *Current Practice of Gas Chromatography-Mass Spectrometry* (pp. 15-43): CRC Press.

Ning, P., Gong, C., Zhang, Y., Guo, K. (2010). La, Ce, Pr, Nd and Sm concentrations in Pu'er tea of Yunnan, China. *Journal of Rare Earths(English Edition), 28*(4), 636-640.

Novak, I., Šeruga, M., Komorsky-Lovrić, Š. (2010). Characterisation of catechins in green and black teas using square-wave voltammetry and RP-HPLC-ECD. *Food Chemistry, 122*(4), 1283-1289.

Okinda Owuor, P., Obanda, M., Nyirenda, H. E., Mphangwe, N. I. K., Wright, L. P., Apostolides, Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (Camellia sinensis) produced in Kenya and comparison with similar teas from Malawi and South Africa. *Food Chemistry, 97*(4), 644-653. doi:10.1016/j.foodchem.2005.04.027

Olafsdottir, G., Martinsdóttir, E., Oehlenschläger, J., Dalgaard, P., Jensen, B., Undeland, I., Mackie, I. M., Henehan, G., Nielsen, J., Nilsen, H. (1997). Methods to evaluate fish freshness in research and industry. *Trends in Food Science & Technology, 8*(8), 258-265.

Owuor, P. O., Obanda, M. (1998). The changes in black tea quality due to variations of plucking standard and fermentation time. *Food Chemistry, 61*(4), 435-441.

Owuor, P. O., Tsushida, T., Horita, H., Murai, T. (1988). Effects of geographical area of production on the composition of the volatile flavour compounds in Kenyan clonal black CTC teas. *Experimental Agriculture, 24*(2), 227-235.

Owuor, P. O. (1992). Comparison of gas chromatographic volatile profiling methods for assessing the flavour quality of Kenyan black teas. *Journal of the Science of Food and Agriculture, 59*(2), 189-197.

Owuor, P. O., Obanda, M., Nyirenda, H. E., Mphangwe, N. I. K., Wright, L. P., Apostolides, Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (Camellia sinensis) produced in Kenya and comparison with similar teas from Malawi and South Africa. *Food Chemistry, 97*(4), 644-653.

Owuor, P. O., Obanda, M. (2007). The use of green tea (Camellia sinensis) leaf flavan-3-ol composition in predicting plain black tea quality potential. *Food Chemistry, 100*(3), 873-884. doi:10.1016/j.foodchem.2005.10.030

Owuor, P. O., McDowell, I. (1994). Changes in theaflavin composition and astringency during black tea fermentation. *Food Chemistry, 51*(3), 251-254. doi:10.1016/0308-8146(94)90023-x

Owuor, P. O., Mang'oka, R., Kavoi, M. M. (2004). Changes in thearubigin fractions and theaflavin levels due to variations in processing conditions and their influence on black tea liquor brightness and total colour.

Palit, M., Tudu, B., Dutta, P. K., Dutta, A., Jana, A., Roy, J. K., Bhattacharyya, N., Bandyopadhyay, R., Chatterjee, A. (2010). Classification of black tea taste and correlation with tea taster's mark using voltammetric electronic tongue. *IEEE transactions on instrumentation and measurement, 59*(8), 2230-2239.

Panigrahi, N., Bhol, C. S., Das, B. S. (2016). Rapid assessment of black tea quality using diffuse reflectance spectroscopy. *Journal of Food Engineering, 190*, 101-108.

Parra, V., Arrieta, Á. A., Fernández-Escudero, J. A., García, H., Apetrei, C., Rodríguez-Méndez, M. L., de Saja, J. A. (2006). E-tongue based on a hybrid array of voltammetric sensors based on phthalocyanines, perylene derivatives and conducting polymers: Discrimination capability towards red wines elaborated with different varieties of grapes. *Sensors and actuators B: Chemical, 115*(1), 54-61.

Pavlou, A. K., Magan, N., Jones, J. M., Brown, J., Klatser, P., Turner, A. P. F. (2004). Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system. *Biosensors and Bioelectronics, 20*(3), 538-544.

Pelillo, M., Bonoli, M., Biguzzi, B., Bendini, A., Toschi, T. G., Lercker, G. (2004). An investigation in the use of HPLC with UV and MS-electrospray detection for the quantification of tea catechins. *Food Chemistry, 87*(3), 465-470.

Poon, G. K. (1998). Analysis of catechins in tea extracts by liquid chromatography–electrospray ionization mass spectrometry. *Journal of Chromatography A, 794*(1-2), 63-74.

Pripdeevech, P., Machan, T. (2011). Fingerprint of volatile flavour constituents and antioxidant activities of teas from Thailand. *Food Chemistry, 125*(2), 797-802.

Punyasiri, P. A. N., Jeganathan, B., Kottawa-Arachchi, J. D., Ranatunga, M. A. B., Abeysinghe, I. S. B., Gunasekare, M. T. K., Bandara, B. M. R. (2017). Genotypic variation in biochemical compounds of the Sri Lankan tea (Camellia sinensis L.) accessions and their relationships to quality and biotic stresses. *The Journal of Horticultural Science and Biotechnology, 92*(5), 502-512.

Rakhmawati, A., Yuliastuti, E. (2011). *Compact computer vision for black tea quality evaluation based on the black tea particles*.

Ren, G., Wang, S., Ning, J., Xu, R., Wang, Y., Xing, Z., Wan, X., Zhang, Z. (2013). Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *Food Research International, 53*(2), 822-826.

Reto, M., Figueira, M. E., Filipe, H. M., Almeida, C. M. M. (2007). Analysis of vitamin K in green tea leafs and infusions by SPME–GC-FID. *Food Chemistry, 100*(1), 405-411.

Röck, F., Barsan, N., Weimar, U. (2008). Electronic nose: current status and future trends. *Chemical Reviews, 108*(2), 705-725.

Rodríguez-Méndez, M. L., De Saja, J. A., González-Antón, R., García-Hernández, C., Medina-Plaza, C., García-Cabezón, C., Martín-Pedrosa, F. (2016). Electronic Noses and Tongues in Wine Industry. *Frontiers in Bioengineering and Biotechnology, 4*. doi:10.3389/fbioe.2016.00081

Roy, R. B., Tudu, B., Shaw, L., Jana, A., Bhattacharyya, N., Bandyopadhyay, R. (2012). Instrumental testing of tea by combining the responses of electronic nose and tongue. *Journal of Food Engineering, 110*(3), 356-363.

Roy, R. B., Modak, A., Mondal, S., Tudu, B., Bandyopadhyay, R., Bhattacharyya, N. (2013). Fusion of electronic nose and tongue response using fuzzy based approach for black tea classification. *Procedia Technology, 10*, 615-622.

Sano, M., Tabata, M., Suzuki, M., Degawa, M., Miyase, T., Maeda-Yamamoto, M. (2001). Simultaneous determination of twelve tea catechins by high-performance liquid chromatography with electrochemical detection. *Analyst, 126*(6), 816-820.

Sapozhnikova, Y. (2014). Development of liquid chromatography–tandem mass spectrometry method for analysis of polyphenolic compounds in liquid samples of grape juice, green tea and coffee. *Food Chemistry, 150*, 87-93.

Scampicchio, M., Benedetti, S., Brunetti, B., Mannino, S. (2006). Amperometric electronic tongue for the evaluation of the tea astringency. *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis, 18*(17), 1643-1648.

Scharbert, S., Hofmann, T. (2005). Molecular definition of black tea taste by means of quantitative studies, taste reconstitution, and omission experiments. *Journal of Agricultural and Food Chemistry, 53*(13), 5377-5384.

Scharbert, S., Holzmann, N., Hofmann, T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *Journal of Agricultural and Food Chemistry, 52*(11), 3498-3508.

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

Sehra, G., Cole, M., Gardner, J. W. (2004). Miniature taste sensing system based on dual SH-SAW sensor device: an electronic tongue. *Sensors and actuators B: Chemical, 103*(1-2), 233-239.

Sharma, M., Thomas, E. V. (2013). Electronic vision study of tea grains color during Infusion. *International Journal of Engineering Science Invention, 2*(4), 52-58.

Sinija, V. R., Mishra, H. N. (2011). Fuzzy analysis of sensory data for quality evaluation and ranking of instant green tea powder and granules. *Food and Bioprocess Technology, 4*(3), 408-416.

Skov, T., Bro, R. (2005). A new approach for modelling sensor based data. *Sensors and actuators B: Chemical, 106*(2), 719-729.

Sliwinska, M., Wisniewska, P., Dymerski, T., Namiesnik, J., Wardencki, W. (2014). Food analysis using artificial senses. *Journal of Agricultural and Food Chemistry, 62*(7), 1423-1448.

Smyth, H., Cozzolino, D. (2012). Instrumental Methods (Spectroscopy, Electronic Nose, and Tongue) As Tools To Predict Taste and Aroma in Beverages: Advantages and Limitations. *Chemical Reviews, 113*(3), 1429-1440. doi:10.1021/cr300076c

Sohn, Y.-S., Goodey, A., Anslyn, E. V., McDevitt, J. T., Shear, J. B., Neikirk, D. P. (2005). A microbead array chemical sensor using capillary-based sample introduction: toward the development of an "electronic tongue". *Biosensors and Bioelectronics, 21*(2), 303-312.

Steinmann, D., Ganzera, M. (2011). Recent advances on HPLC/MS in medicinal plant analysis. *Journal of Pharmaceutical and Biomedical Analysis, 55*(4), 744-757.

Stone, H., Sidel, J. L. (2004). Sensory Evaluation Practices, Elsevier Academic Press. *California, USA*.

Strike, D. J., Meijerink, M. G. H., Koudelka-Hep, M. (1999). Electronic noses–A mini-review. *Fresenius' journal of analytical chemistry, 364*(6), 499-505.

Sun, D.-W. (2016). *Computer vision technology for food quality evaluation*: Academic Press.

Tahara, Y., Toko, K. (2013). Electronic tongues–a review. *IEEE Sensors Journal, 13*(8), 3001-3011.

Tao, W., Zhou, Z., Zhao, B., Wei, T. (2016). Simultaneous determination of eight catechins and four theaflavins in green, black and oolong tea using new HPLC-MS-MS method. *J Pharm Biomed Anal, 131*, 140-145. doi:10.1016/j.jpba.2016.08.020

Tarachiwin, L., Ute, K., Kobayashi, A., Fukusaki, E. (2007). 1H NMR based metabolic profiling in the evaluation of Japanese green tea quality. *Journal of Agricultural and Food Chemistry, 55*(23), 9330-9336.

Tozlu, B. H., Okumuş, H. İ. (2018). A new approach to automation of black tea fermentation process with electronic nose. *Automatika, 59*(3-4), 373-381.

Tu, Y., Bian, M., Wan, Y., Fei, T. (2018). Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. *PeerJ, 6*, e4858.

Tudu, B., Jana, A., Metla, A., Ghosh, D., Bhattacharyya, N., Bandyopadhyay, R. (2009). Electronic nose for black tea quality evaluation by an incremental RBF network. *Sensors and actuators B: Chemical, 138*(1), 90-95.

Vasefi, F., MacKinnon, N., Farkas, D. L. (2016). Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology. In M. R. Hamblin, P. AvciG. K. Gupta (Eds.), *Imaging in Dermatology* (pp. 187-201). Boston: Academic Press.

Venzie, J. L., Castro, J., Krishna, M. V. B., Nelson, D. M., Marcus, R. K. (2007). Electron-impact and glow-discharge ionization LC–MS analysis of green tea tincture. *Analytical and bioanalytical chemistry, 387*(1), 321-333.

Vo-Dinh, T. (2004). A hyperspectral imaging system for in vivo optical diagnostics. *IEEE Engineering in Medicine and Biology Magazine, 23*(5), 40-49.

Wang, D., Lu, J., Miao, A., Xie, Z., Yang, D. (2008). HPLC-DAD-ESI-MS/MS analysis of polyphenols and purine alkaloids in leaves of 22 tea cultivars in China. *Journal of Food Composition and Analysis, 21*(5), 361-369.

Wang, K., Ruan, J. (2009). Analysis of chemical components in green tea in relation with perceived quality, a case study with Longjing teas. *International Journal of Food Science & Technology, 44*(12), 2476-2484.

Wang, K., Liu, F., Liu, Z., Huang, J., Xu, Z., Li, Y., Chen, J., Gong, Y., Yang, X. (2010). Analysis of chemical components in oolong tea in relation to perceived quality. *International Journal of Food Science & Technology, 45*(5), 913-920. doi:10.1111/j.1365-2621.2010.02224.x

Wang, L.-F., Lee, J.-Y., Chung, J.-O., Baik, J.-H., So, S., Park, S.-K. (2008). Discrimination of teas with different degrees of fermentation by SPME–GC analysis of the characteristic volatile flavour compounds. *Food Chemistry, 109*(1), 196-206.

Wright, L. P. (2002). Biochemical analysis for identification of quality in black tea (Camellia sinensis).

Wu, F., He, Z., Luo, Q. (1998). High-performance liquid chromatographic determination of oxalic acid in tea using tris (1, 10-phenanthroline)-ruthenium (II) chemiluminescence. *Analytical Sciences, 14*(5), 971-973.

Wu, J., Liu, J., Fu, M., Li, G. (2006). Classification of Chinese green tea by a voltammetric electronic tongue. *Chinese Journal of Sensors and Actuators, 4*, 007.

Wu, R., Zhao, J., Chen, Q., Huang, X. (2011). Quality assessment of green tea taste by using electronic tongue. *Transactions of the Chinese Society of Agricultural Engineering, 27*(11), 378-381.

Xiao, H., Wang, J. (2009). Discrimination of Xihulongjing tea grade using an electronic tongue. *African Journal of Biotechnology, 8*(24).

Xiong, C., Liu, C., Pan, W., Ma, F., Xiong, C., Qi, L., Chen, F., Lu, X., Yang, J., Zheng, L. (2015). Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system. *Food Chemistry, 176*, 130-136.

Xu, L., Deng, D.-H., Cai, C.-B. (2011). Predicting the Age and Type of Tuocha Tea by Fourier Transform Infrared Spectroscopy and Chemometric Data Analysis. *Journal of Agricultural and Food Chemistry, 59*(19), 10461-10469. doi:10.1021/jf2026499

Xu, N., Chen, Z. M. (2002). Green tea, black tea and semi-fermented tea. *Tea: Bioactivity and Therapeutic Potential*, 35-57.

Xu, Y.-Q., Wang, C., Li, C.-W., Liu, S.-H., Zhang, C.-X., Li, L.-W., Jiang, D.-H. (2016). Characterization of aroma-active compounds of pu-erh tea by headspace solid-phase microextraction (HS-SPME) and simultaneous distillation-extraction (SDE) coupled with GC-olfactometry and GC-MS. *Food Analytical Methods, 9*(5), 1188-1198.

Yamamoto, T., Juneja, L. R., Chu, D. C., Kim, M. (1997). Chemistry and Applications of Green Tea CRC press. *Florida, USA*.

Yamanishi, T. (1995). Special issue on tea: Flavor of tea. *Food Review Int., 11*, 477-525.

Yamauchi, Y., Nakamura, A., Kohno, I., Kitai, M., Hatanaka, K., Tanimoto, T. (2008). Simple and rapid UV spectrophotometry of caffeine in tea coupled with sample pre-treatment using a cartridge column filled with polyvinylpolypyrrolidone (PVPP). *Chemical and Pharmaceutical Bulletin, 56*(2), 185-188.

Yang, X. R., Ye, C. X., Xu, J. K., Jiang, Y. M. (2007). Simultaneous analysis of purine alkaloids and catechins in Camellia sinensis, Camellia ptilophylla and Camellia assamica var. kucha by HPLC. *Food Chemistry, 100*(3), 1132-1136.

Yang, Y.-Q., Yin, H.-X., Yuan, H.-B., Jiang, Y.-W., Dong, C.-W., Deng, Y.-L. (2018). Characterization of the volatile components in green tea by IRAE-HS-SPME/GC-MS combined with multivariate analysis. *PLoS One, 13*(3), e0193393.

Yashin, A. Y., Nemzer, B. V., Combet, E., Yashin, Y. I. (2015). Determination of the chemical composition of tea by chromatographic methods: a review. *Journal of Food Research, 4*(3), 56.

Ye, J., Wang, W., Ho, C., Li, J., Guo, X., Zhao, M., Jiang, Y., Tu, P. (2016). Differentiation of two types of pu-erh teas by using an electronic nose and ultrasound-assisted extraction-dispersive liquid–liquid microextraction-gas chromatography-mass spectrometry. *Analytical Methods, 8*(3), 593-604.

Ye, N., Zhang, L., Gu, X. (2012). Discrimination of green teas from different geographical origins by using HS-SPME/GC–MS and pattern recognition methods. *Food Analytical Methods, 5*(4), 856-860.

Yu, H., Wang, J., Xiao, H., Liu, M. (2009). Quality grade identification of green tea using the eigenvalues of PCA based on the E-nose signals. *Sensors and actuators B: Chemical, 140*(2), 378-382.

Yu, H., Wang, J. (2007). Discrimination of LongJing green-tea grade by electronic nose. *Sensors and actuators B: Chemical, 122*(1), 134-140.

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 Science and Technology, 41*(7), 1268-1273.

Yu, P., Yeo, A. S., Low, M. Y., Zhou, W. (2014). Identifying key non-volatile compounds in ready-to-drink green tea and their impact on taste profile. *Food Chem, 155*, 9-16. doi:10.1016/j.foodchem.2014.01.046

Yuan, Y., Song, Y., Jing, W., Wang, Y., Yang, X., Liu, D. (2014). Simultaneous determination of caffeine, gallic acid, theanine,(−)-epigallocatechin and (−)-epigallocatechin-3-gallate in green tea using quantitative 1 H-NMR spectroscopy. *Analytical Methods, 6*(3), 907-914.

Zampetti, E., Pantalei, S., Macagnano, A., Proietti, E., Di Natale, C., D'amico, A. (2008). Use of a multiplexed oscillator in a miniaturized electronic nose based on a multichannel quartz crystal microbalance. *Sensors and actuators B: Chemical, 131*(1), 159-166.

Zareef, M., Chen, Q., Ouyang, Q., Kutsanedzie, F. Y. H., Hassan, M. M., Viswadevarayalu, A., Wang, A. (2018). Prediction of amino acids, caffeine, theaflavins and water extract in black tea using FT-NIR spectroscopy coupled chemometrics algorithms. *Analytical Methods, 10*(25), 3023-3031. doi:10.1039/C8AY00731D

Zeeb, D. J., Nelson, B. C., Albert, K., Dalluge, J. J. (2000). Separation and identification of twelve catechins in tea using liquid chromatography/atmospheric pressure chemical ionization-mass spectrometry. *Analytical Chemistry, 72*(20), 5020-5026.

Zheng, X.-Q., Li, Q.-S., Xiang, L.-P., Liang, Y.-R. (2016a). Recent advances in volatiles of teas. *Molecules, 21*(3), 338.

Zheng, X.-Q., Li, Q.-S., Xiang, L.-P., Liang, Y.-R. (2016b). Recent Advances in Volatiles of Teas. *Molecules, 21*(3). doi:10.3390/molecules21030338

Zhi, R., Zhao, L., Shi, B., Jin, Y. (2014). New dimensionality reduction model (manifold learning) coupled with electronic tongue for green tea grade identification. *European Food Research and Technology, 239*(1), 157-167.

Zhi, R., Zhao, L., Zhang, D. (2017). A framework for the multi-level fusion of electronic nose and electronic tongue for tea quality assessment. *Sensors, 17*(5), 1007.

Zhu, H., Ye, Y., He, H., Dong, C. (2017). Evaluation of green tea sensory quality via process characteristics and image information. *Food and Bioproducts Processing, 102*, 116-122. doi:10.1016/j.fbp.2016.12.004

Zhuang, X., Wang, L., Chen, Q., Wu, X., Fang, J. (2017). Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools. *Science China Technological Sciences, 60*(1), 84-90. doi:10.1007/s11431-016-0464-0

Zoecklein, B. W., Devarajan, Y. S., Mallikarjunan, K., Gardner, D. M. (2011). Monitoring effects of ethanol spray on Cabernet franc and Merlot grapes and wine volatiles using electronic nose systems. *American journal of enology and viticulture*, ajev-2011.

Zou, G., Xiao, Y., Wang, M., Zhang, H. (2018). Detection of bitterness and astringency of green tea with different taste by electronic nose and tongue. *PLoS One, 13*(12), e0206517.

Evaluation of an optimized miniature process for use in black tea quality assessment in a breeding programme

Lubanga M. Nelson[b,c*], De-Silva Jacquie[d], Massawe Festo[b], Mayes Sean[a]

*nelsonlubanga@gmail.com

[a]School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK.

[b]School of biosciences, University of Nottingham, Malaysia Campus Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[c]Unilever Tea Kenya Ltd., Nakuru Road, Kericho, 20-20200, Kenya.

[d]Unilever Research & Development, Colworth Science Park, Sharnbrook, Bedford MK441LQ, UK.

Abstract

Tea quality is an important, yet a complex attribute that is influenced by the manufacturing process, variety, environmental conditions and post-harvest handling. Manufacturing of different tea varieties under varying environmental conditions is a challenging task. The objective of this research was to investigate the reliability of using an optimized miniature process for manufacturing newly developed varieties into black tea for assessment of their quality potential before release for commercial planting. Four commercially grown Kenyan tea varieties were manufactured into black tea using an optimised miniature process. Objective colour measurements, total soluble solids, total polyphenol content and taster scores were measured. Significant differences were observed for all the objective measurements except infusion b*. Results of principle component analysis (PCA) and hierarchical cluster analysis (HCA) indicated differences in the patterns of the four varieties suggesting that the new optimized miniature process was reproducible, although still with some variation between replicates. Among the 4 varieties evaluated, CLONE8was distinguished from the rest. However, sensory evaluation results did not differentiate the varieties, but indicated their quality was high. Many varieties are developed in a tea breeding programme, each with unique quality properties. Although, the present optimized miniature process can be used for

manufacturing different tea varieties, technologies that can help in achieving optimum withering and fermentation conditions for many tea varieties under different environmental conditions could be explored. Objective methods of evaluating polyphenols, caffeine, amino acids and aroma compounds present in fresh tea shoots could also be considered and used in predicting high quality black tea for the newly developed varieties.

Key words: Black tea quality; miniature manufacture; tea breeding programme

5.1. Introduction

Tea quality is the main factor influencing price at the tea auction (Wei et al., 2011) and hence it is an important attribute for tea companies and consumers (Mondal, 2014a). Miniature manufacture of tea involves processing of tea in small quantities (2 – 5 kilograms of leaf samples). It is an important step in a tea breeding programme and is used to process tea varieties that are still in trials to evaluate their black tea quality before release for commercial planting. Miniature manufacturing of tea simulates the normal factory process; withering (loss of moisture); maceration (softening of the leaf); cutting (breaking down the cellular matrix and structure); fermentation (aeration); drying and sorting (Tang et al., 2018). A reliable miniature process is highly desirable for processing many varieties under trial, to enable evaluation of their black tea quality properties.

The current miniature process used to manufacture different samples is not reproducible. The success of tea manufacturing is dependent on the genotype, environmental conditions (weather conditions, soil types, season), agronomic factors (cultivation, type and quantity of fertilizer applied, plucking method, type of leaf harvested, green leaf handling), manufacturing process and storage of made tea (Zheng et al., 2016a). Different tea varieties have unique biochemical characteristics (Ho et al., 2015) and hence the processing conditions for each clone is different. Environmental conditions in the field such as rainfall, temperature and humidity fluctuate most of the time, and the processing conditions need to be optimised to achieve high quality made tea. Similarly, the quality of the harvested shoots and handling significantly influences the quality of made tea. In a breeding programme, many new tea varieties are evaluated, under

varying environmental and post-harvest conditions that are difficult to control. The objective of this research was to investigate the utilization, reproducibility and the major challenges of optimizing a miniature process for manufacturing different tea samples and to propose methods that could be considered for evaluating tea quality.

5.2. Materials and methods

5.2.1. Plant materials and experimental design

The plant material consisted of four popular commercial tea clones; CLONE95 is a known high-quality variety, CLONE8is a high yielding variety, CLONE37 is a high yielding and good quality clone while CLONE101 is a high-quality clone. CLONE95 and CLONE101 are high quality standards (Gabriel Tuwei, Pers. com, June 2016, R&D Unilever Tea Kenya). CLONE8and CLONE37 are newly developed clones and were gazetted in 2011 by the Kenyan government. The four clones were obtained from clonal field trials (CFT) at Jamji estate (0^0 28' S and 35^0 11'E), located 1733 meters above sea level in Kericho county. Each of the 4 clones was harvested, every day for 5 days between 8am – 9am. Variabilities were minimised by plucking two leaves with an apical bud and handling and transporting in the same way and time prior to processing. The 5 biological replicates for each of the four clones were then manufactured separately and subjected to quality measurements separately using objective and subjective methods.

5.2.2. Optimised miniature manufacture process of black tea

A 2 Kg of fresh shoots comprising of two leaves and a bud were harvested in June 2017 from the clonal field trials (CFT) at Unilever Tea Kenya. The plucked fresh tea shoots were loaded in a withering trough at 10 am and ambient air was passed through the leaves for 24 hours using a fan so that physical and chemical withering could take place at room temperature. The withered leaves with between 67% to 72% moisture content (mc) were cut 5 times with ball-breaking after the fifth cut using the Unilever Tea Kenya R&D miniature crush, tear and curl (CTC) machine for 5 minutes (1 minute for each cut) for each sample.

The crushed leaves (dhool) were fermented in an environmental cabinet (Gallenkamp cabinet) for 120 minutes at a set temperature of 32^0C for 15-20 minutes for the first phase and 26^0C during the second phase up to 120^{th} minute. Fermentation was done by turning the dhool after every 30 minutes.

During fermentation, air was blown gently into the environmental cabinet using a small fan. The fermented dhool was dried using a miniature fluidized bed drier. The fibres were removed from the hot dry made tea using an electrostatic fibre board. The dried leaf was then passed through different meshes (1250 µm, 1000 µm, 710 µm, 500 µm, pan) to isolate the different grades of tea. PF1 was isolated at mesh 710 µm and the samples were packed into sealed aluminium sachets for analysis.

5.2.3. Colour measurements

Colour measurement of brewed tea liquor was carried out using a chroma minolta meter (Model CT 310) and values were calculated by the CIE system (Commission Internationale de l'Éclairage). Minolta chroma meter is a compact tristimulus colour analyser for measuring reflective colours of surfaces objectively. Each of the 5 biological replicates for each clone was measured and the parameters evaluated were infusion L, a and b values. Colour measurement of dry made tea was done using a DigiEye system to determine L*, a*, b*, chroma, hue angle and saturation for the 5 biological replicates of each of the four clones. Samples were placed into a small petri dish, put into the DigiEye system and measurements taken.

5.2.4. Determination of total soluble solids (TSS)

A 200 ml of boiling mineral water was added into a plastic beaker containing 2 g of black tea. The mixture was allowed to brew for 2 minutes and then filtered through two layers of muslin. The infusion then cooled down to room temperature. A 100 ml foil tray was weighed and recorded (m1). 50 ml of tea infusion was pipetted into the tray and weighed (m2). The foil tray was then placed in an oven at 100^0C overnight and put to a desiccator to cool before weighing (m3).

Total soluble solids were calculated using the following equation:

%SS = (m3 – m1)/ (m2 – m1) x 100

Where, % SS = % Soluble solids in infusion, m1= Weight empty Tray, m2 = Weight tray + Tea infusion and m3 = Weight tray + Dry tea solids.

5.2.5. Total Polyphenols analysis in black tea

The total polyphenol content was determined using a spectrophotometer by following the procedure from ISO 14502-1:2005. Gallic acid equivalents in 0.5 g of tea powder was collected from each of the 20 samples and determined against a standard curve generated using gallic acid (Figure S5.1). The total polyphenol content was expressed as a percentage by mass on a sample dry matter basis. 1 ml of prepared gallic acid standard solutions (A to E) was measured using a pipette and transferred in duplicate into separate tubes which responded to 10, 20, 30, 40 and 50 µg gallic acid standards. 1 ml of water (blank solution) was measured in duplicate and transferred into separate disposable tubes. 1 ml of the sample was measured in duplicate into separate disposable tubes. 5 ml of Folin & Ciocalteu working reagent was added into each tube. After 3 to 8 minutes, 4 ml of 7.5% sodium carbonate solution was added into each tube. The reagents were stoppered, mixed and left to stand at room temperature for 1 hour. The absorbance was measured at 750 nm in a 1 cm cell.

5.2.6. Organoleptic tea quality evaluation

Quality scores were blindly assessed by a trained panel of three professional tasters at the Mombasa tea auction in Kenya. All the processed teas were first infused as follows; 5.6 grams of the sample was measured using a hand scale and transferred into an infusion pot. De-ionized water was boiled and added into the infusion pot until it was full. The boiled water and sample were left to brew for 6 minutes. Using a syringe, 5 ml of milk was carefully placed into each tasting bowl. The infused liquor was then filtered into the tasting bowl containing milk and left to stand for 10 minutes before it was analysed. The residue from the infusion was collected on the infusion cup lid. The liquor was then analysed accordingly.

5.2.7. Statistical analysis

Analysis of variance and standard deviations for all traits were done using the GenStat statistical software, 18th Edition (Payne et al., 2009). Significant differences among sample means were compared using Duncan's multiple range test at $p<0.05$ level of significance. Data was expressed as means ± SD of 5 biological replicates with $P < 0.05$ representing a statistically significant difference.

Multivariate analysis using principal component analysis (PCA) and hierarchical cluster analysis (HCA) of the traits under study was performed using Qlucore omics Explorer 3.4, to explore the clustering of the tea samples based on the genotype. A Qlucore compatible table of the quality measurements (variables) and samples with 3 annotations (sample code, batch and clone) was created in excel and saved as a tab delimited file with a gedata suffix on a PC with the Qlucore Omics Explorer 3.4 software and loaded. A total of 20 entries comprising of 4 clones replicated 5 times were analysed. A sample by trait dimensional matrix of the data set was displayed on the heatmap where each column represents the standardized binary measurements for a given sample. Higher values for a particular attribute is coded as red and lower values coded as green. The intensities of red and green colours represent the proportion of more and less respectively of a particular attribute for a given clone.

5.3. Results

5.3.1. Colour of dry made tea

Differences in colour of made tea leaf using the DigiEye were observed in the four clones, and the results are summarized in Table 1. There were significant differences ($p<0.05$) for all the measured traits, indicating variations in colour of dry tea among the four clones (Table 5).

Table 5. Mean colour measurement values for dry leaf obtained from DigiEye. Values showing *, ** and * stand for significance at 0.05, 0.01 and 0.001 probability level, respectively.**

Clone	L*	a*	b*	chroma	Hue angle (^{0})	Saturation
CLONE8	18.01	7.79	14.66	17.08	55.66	42.82

CLONE37	17.51	8.74	12.23	16.09	48.08	41.83
CLONE95	16.50	7.70	9.52	13.49	44.7	38.78
CLONE101	17.29	8.5	11.06	15.01	46.18	40.65
CV (%)	3	4.3	13.8	6.9	8.2	2.8
MS	2.91**	1.25***	31.49**	21.83***	163.31**	17.71***
SD	0.819	0.65	2.41	1.76	5.42	1.94

The L* a* b* measurements were used to define dry leaf colour properties of the four varieties. L* a* b* are colour space as defined by the CIE based on luminance (lightness) (L*) and a* and b* values. L* is the lightness co-ordinate and has values ranging from 0 to 100, whereby 0 represents black while 100 characterizes white. Therefore, lower values of L* are desirable in black tea quality evaluation. L* values ranged from 16.50 (CLONE95) to 18.01 (CLONE8) respectively (Table 5), implying that CLONE95 is darker than the other 3 clones and is of higher black colour quality. Redness and greenness are described by a* values, and positive values indicate redness and negative values indicate greenness. The a* values ranged from 7.70 (CLONE95) to 8.74 (CLONE37) (Table 5), implying that CLONE95 has less redness while CLONE37 is the most red in colour. However, the range is small and not likely to be noticeable. Positive b* values indicate more yellow colour, while negative values indicate more blue colour. In this study, b* values ranged from 9.52 (CLONE95) to 14.66 (CLONE8) (Table 5). Clone CLONE8had significantly ($p<0.05$) higher b* values compared to the other three clones, implying that it has a more yellowish leaf appearance.

Hue angle (h*) is a qualitative attribute where colours are described as reddish, greenish, yellowish and bluish. It is used to define the differences in colours with reference to grey with the same lightness. An angle of 0° or 360° characterizes red hue, whilst angles of 90°, 180° and 270° describe yellow, green and blue hues respectively. In this study, hue angle ranged from 44.70 (CLONE8) to 55.70 (CLONE95), suggesting that all the four clones sit within the red region of hue. CLONE8had a significantly ($p<0.05$) greater hue angle compared to the rest of the clones, suggesting that it has a more yellow hues than the other 3 varieties.

Chroma is a quantitative attribute of colourfulness used to determine the degree of the differences of hue in comparison to a grey colour with the same lightness. The higher the chroma values, the higher is the colour intensity as perceived by humans. It has an open-ended scale with a zero origin representing neutrals with no hue. Chroma ranged from 13.49 (CLONE95) to 17.08 (CLONE8) (Table 5), implying that CLONE8had a greater colour intensity than the other clones. Saturation values reflected those of Chroma and ranged from 38.78 (CLONE95) to 42.82 (CLONE8) (Table 5), showing CLONE8was the brightest. Saturation refers to the intensity of colour in an image.

5.3.2. Measurement of traits relating to infused tea

There were significant clonal differences ($p<0.05$) for Infusion L* and Infusion a*, indicating variations in colour of the infused tea among the four clones (Table 6).

Table 6. Mean L*a*b*, Total Solids, Total Polyphenols and sensory evaluation values. Values showing *, ** and * stand for significance at 0.05, 0.01 and 0.001 probability level, respectively. ns is non-significant difference ($p<0.05$).**

Clone	Infusion L*	Infusion a*	Infusion b*	TS	TPP	Leaf	Taste	Mouth feel	Hue
CLONE8	66.94	24.44	100.91	24.62	673.46	5	5	4.6	5.28
CLONE37	64.01	26.97	100.73	24.78	676.40	5	5	4.6	5.04
CLONE95	64.10	27.34	101.31	23.59	574.52	5	5	4.6	5.16
CLONE101	62.56	29.05	101.67	24.98	695.13	5	5	4.6	5.20
CV (%)	3.2	8.1	1.4	2.9	4.8				
LSD (P<0.05)	2.82	3.02	1.92	0.98	43.26				
MS	17.39*	20.195*	1.183 ns	2.117*	15361***				
SD	2.37	2.57	1.37	1.25	58.08				

Infusion L*, a*, b* values were obtained using a minolta chroma meter after infusing the made tea samples. Infusion L* values ranged from 62.56 (CLONE101) to 66.94 (CLONE8) (Table 6). CLONE8had significantly ($p<0.05$) the highest mean infusion L* values compared to the other clones, implying that it has lighter liquors. Infusion a* values ranged from 24.44 (CLONE8) to 29.05 (CLONE101) (Table 6), suggesting that CLONE101 liquor is redder than CLONE8. Infusion b* values ranged from 100.73

(CLONE37) to 101.7 (CLONE101) (Table 6). However, infusion b* values were not significantly ($p<0.05$) different for all the samples.

Total soluble components in tea influences the development of the flavour in tea (Someswararao et al., 2013). Total soluble solids ranged from 23.59 (CLONE95) to 24.98 (Clone101) (Table 6). Clone95 had significantly ($p<0.05$) the lower total soluble solids compared to the other three clones. However, the levels of total soluble solids in CLONE101, CLONE37 and CLONE8 were not significantly ($p<0.05$) different from each other.

Theaflavins and thearubigins, oxidised products of polyphenols are responsible for the taste and colour of black tea (Owuor et al., 2007; Scharbert et al., 2004; Obanda et al., 1997). There were significant ($p<0.05$) variations in the composition of total polyphenols among the clones, indicating variations in the quality of the four clones. Total polyphenol content ranged from 574.52 (CLONE95) to 695.1 (CLONE101) (Table 6). The levels of total polyphenols in CLONE95 was significantly ($p<0.05$) lower compared to the other three clones.

5.3.3. Organoleptic tea quality evaluation

Organoleptic tea quality evaluation involves professional tea tasters who assess tea for desirable attributes such as mouthfeel, flavour and colour of tea (Hazarika, 2012). Taste is described by astringency, bitterness, mellowness and slight sweetness (Bhuyan et al., 2009). Mouth feel is the heaviness, thickness and strength of tea liquor while hue is the colour of tea after infusion and it ranges from pale yellow to dark red. Leaf is the appearance of dry tea after manufacture. Organoleptic evaluation results for the four varieties are presented in Table 2. Taster scores results showed that taste, mouthfeel and colour of dry made tea and infusion were not significantly ($p<0.05$) different among the four clones, suggesting that the tea tasters could not differentiate the clones. However, all the varieties were scored highly on taste, and colour of dry tea (leaf) and infusion (hue) (Table 6), implying that the mini-manufacturing process produces high quality teas. Since all the teas are commercially grown and of acceptable quality, it is evident that the optimised miniature process is reflective of the factory process.

5.3.4. Multivariate analysis

5.3.4.1. Principle component analysis (PCA)

PCA was used to investigate differences among the four clones. A PCA plot was generated from all the traits investigated with samples coloured by clone. The PCA plot explained 82% of the existing variation in the four varieties, suggesting considerable diversity among the genotypes. From the PCA plot, the samples were successfully divided into three distinct groups, which is an indication of differences existing between the varieties based on their quality properties. Principal component 1 explained the greatest variation at 47%, while principal component 2 and principal component 3 explained the least variation at 23% and 12%, respectively (Figure 1). Clone8 and CLONE95 were separated into 2 distinct groups while Clone101 and Clone37 were in the same cluster (Figure 5). This suggests that this miniature process could produce black tea reproducibly and can also allow discrimination of clones significantly different from each other.

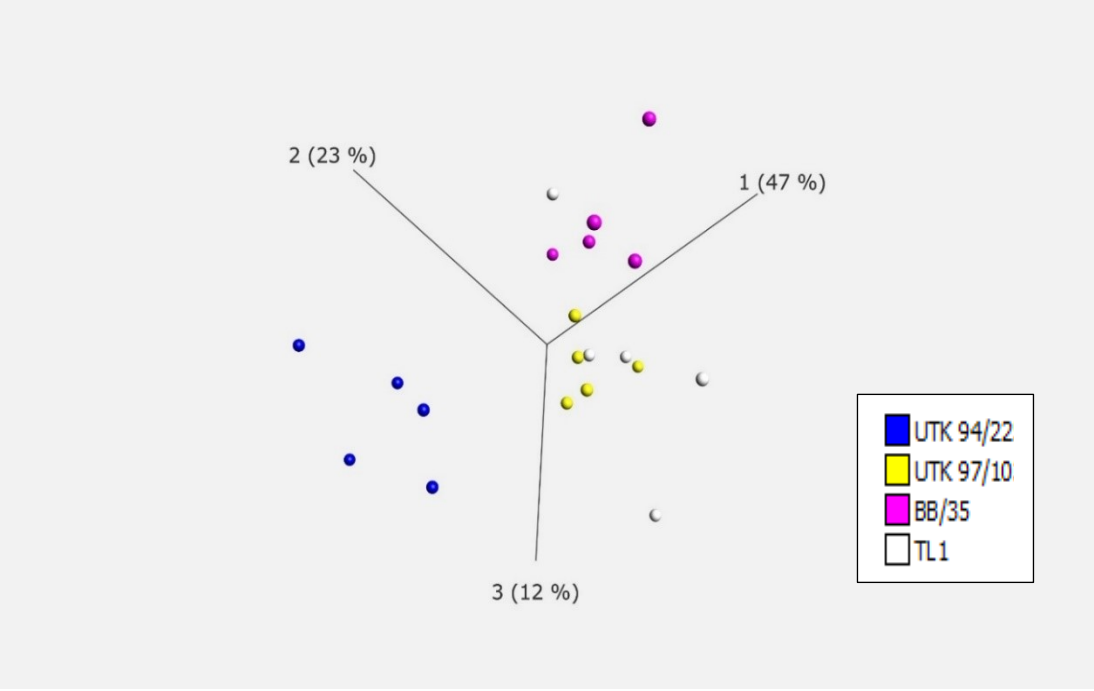

Figure 5. A PCA plot was generated with samples coloured by clone.

5.3.4.2. Hierarchical cluster analysis (HCA)

A heat map was generated with variables ordered hierarchically by characteristic (shared profile across all clones) and by clone (shared profile across all characteristics). The HCA dendrogram (Figure 6) shows that samples were separated into three groups, suggesting that there were differences among the varieties. Clone CLONE8and CLONE95 were clearly differentiated, while Clone101 and CLONE37 overlapped to form another group (Figure 6). It could also imply that Clone101 and CLONE37 could have the same quality characteristics.

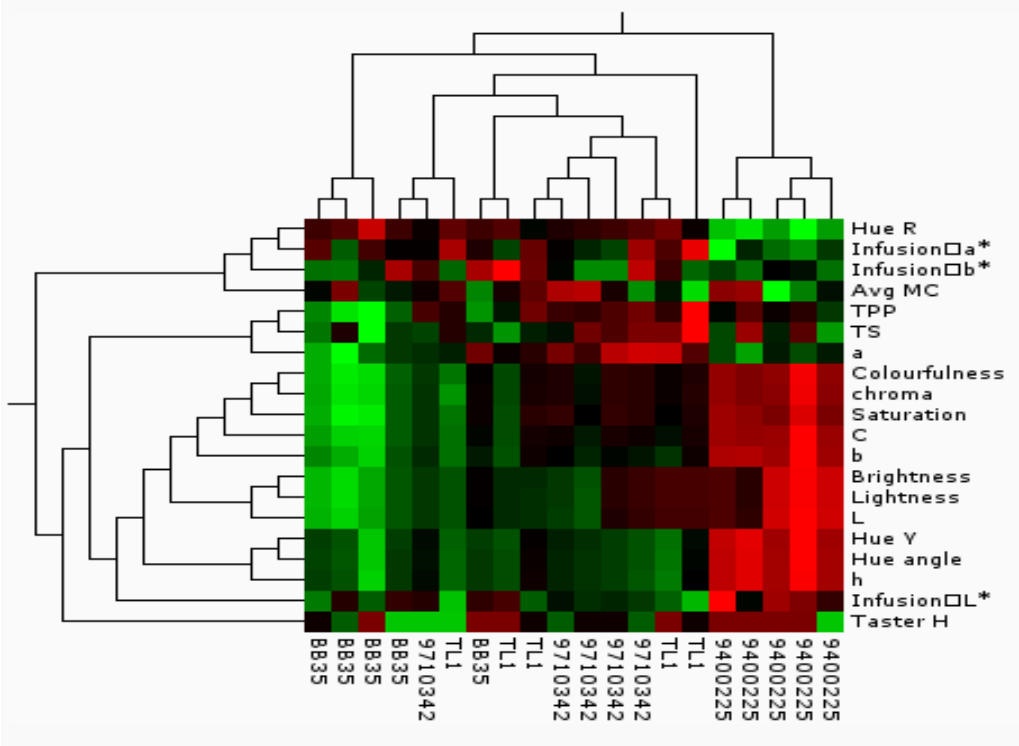

Figure 6. HCA dendrogram of the 4 clones and all the measured traits.

5.4. Discussion

In this study, four commercial varieties were processed using a miniature manufacture equipped and their quality evaluated using objective and subjective measurements. The objective quality evaluation methods involved measuring the colour of infused tea using chroma minolta meter, colour of dry made tea using a DigiEye, total soluble solids and total polyphenols using a spectrophotometer. Results from PCA and HCA showed that the four varieties were differentiated, especially based on the objective tea quality

evaluation measurements. Particularly, CLONE8was clearly distinguished from CLONE95, implying that the two varieties have different quality attributes. The fact that the tasters could not distinguish any of the biological replicates was not surprising, despite scoring the teas as high quality. Most of the tasters are not trained to identify quality differences between different varieties. In most cases, the tea tasters give high scores for well manufactured teas. Therefore, relying on tea tasters alone may result in inaccurate evaluation of tea varieties based on their quality (Chen et al., 2015). This coupled with the fact that sensory evaluation is a highly subjective method, and different tasters may have different quality preferences for the same tea.

Tea manufacture is a complicated process involving withering, maceration, cutting, fermentation, drying and sorting (Tang et al., 2018). Withering (chemical and physical) is an important process in tea manufacture since it contributes to taste and aroma of black tea (Wang et al., 2018a), and therefore it should be monitored carefully. Withering has a significant impact on the subsequent stages of tea processing including maceration, cutting, fermentation, drying and storage (Deb et al., 2016). Chemical withering starts immediately the leaf is plucked and it involves biochemical and physiological changes in the leaf (Han et al., 2016). Physical withering leads to moisture loss and the turgid shoots become flaccid, making it easy to macerate (Owuor et al., 2017). Withering is a slow and difficult process to control (Owuor et al., 2017), hence achieving an optimised physical and chemical wither is the greatest challenge in black tea processing (Deb et al., 2016; Owuor et al., 1997). Withering is influenced by the plucking standards, handling of the harvested leaf, transportation and environmental conditions (rainfall, wind, temperature, soil moisture and humidity) (Deb et al., 2016). For instance, during the rainy season, the tea plant accumulates a lot of water in the green leaf, which requires a long withering duration. On the contrary, during the dry period and on windy days, there are low levels of moisture in the leaves, which requires a shorter withering time. It has been reported that long chemical and hard physical withering durations reduce the quality of black tea, which varies with the genetic makeup of the tea (Owuor et al., 2017). Technologies to help in achieving an optimised wither for the clones under evaluation in a breeding programme should be put in place. The ideal situation in a tea breeding programme is to subject the newly developed clones to a uniform withering

process. However, this is not practical because of the fluctuating environmental conditions and differences in biochemical properties of varieties.

Fermentation process parameters such as time , temperature , oxygen , relative humidity and pH , are the critical factors that affect the quality of black tea and therefore should be optimised (Obanda et al., 2001). Black tea is formed through an enzymatic process that involves oxidation of polyphenols in fresh green tea into theaflavin and thearubigins. Theaflavins contribute to the taste and colour of black tea, while thearubigins contribute to the mouth feel (thickness) and colour of the tea. Fermentation process parameters should be optimised depending on the type of tea variety. Different varieties require a varying degree of fermentation (time, temperature and relative humidity).

Additionally, post-harvest conditions such as plucking standards of tea leaves influence the quality of black tea. It was reported that theaflavins levels, colour and tasters' preferences decreased with coarse plucking (Owuor et al., 1987). Similarly, storage conditions affect the quality of made tea (Zheng et al., 2016a).

At present, testing for black tea quality involves subjecting the newly developed varieties to a standard manufacturing process before analysis. However, each step of tea manufacture is a source of variation. Several studies have linked the quality of black tea with green leaf polyphenols (catechins), caffeine, amino acids and aroma (Zheng et al., 2016a; Ho et al., 2015; Owuor et al., 2007; Liang et al., 2003). Several studies have proposed using green leaf biochemical composition such as catechins and caffeine in selecting clones with high quality black tea early (Wright et al., 2002; Obanda et al., 1997). This makes it advantageous because tea breeding takes a very long time.

Organoleptic evaluation is currently the main method used to determine tea quality using professional tasters, who describe the colour, aroma and taste properties of tea (Hazarika, 2012; Obanda et al., 2004; Liang et al., 2003). However, sensory evaluation is a time-consuming and laborious exercise, which may affect the objectivity of the tasters, and hence may result in inaccurate evaluation of tea quality (Bhattacharyya et al., 2008). This makes organoleptic evaluation to be a highly subjective method (Yu et al., 2014; Liang et al., 2008).

Objective physical and chemical methods of tea quality assessment have been developed and they exploit technologies such as nuclear magnetic resonance spectroscopy (NMR) (Le Gall et al., 2004), high performance liquid chromatography (HPLC) (Zhang et al., 2017c; Yashin et al., 2015), mass spectroscopy (Fraser et al., 2013), near infrared (NIR) spectroscopy (Mishra et al., 2018b), electronic tongue (Yaroshenko et al., 2014), electronic nose (Sharma et al., 2015; Roy et al., 2012), computer vision (Gill et al., 2011a), capillary electrophoresis (Ning et al., 2010) and integrated multi-sensor system (Roy et al., 2012). These methods are accurate and reproducible, and they have been used for discriminating varieties with different quality properties (Zhu et al., 2017).

5.5. Conclusion

Tea is grown outside in the field and the varying environmental conditions (humidity, rainfall, temperature, wind), post-harvest handling (harvested shoots quality standards and time taken to transport to the factory), and different tea varieties significantly affect the processing parameters, hence require frequent modification. More effort should therefore be put to monitoring to ensure that all the tea process parameters are well optimised and fall within the required range. Tea varieties have different genetic make-ups and therefore each require unique process requirements. The ideal situation is to optimize a process for each clone, before evaluating quality. However, this is not practical in a tea breeding programme or factory setting since many clones are developed and harvested, respectively.

The mini manufacture process distinguished the four tea varieties based on their individual biochemical properties as revealed by the PCA, HCA and ANOVA. Clone CLONE8was clearly distinguished from the other three clones for hue angle, hue Y, hue R, leaf b* and Infusion L*. Clone CLONE95 was discriminated from the other clones for brightness, lightness and leaf L*. Clone CLONE101 was also distinguished from the other clones for total soluble solids and total polyphenol content. The professional tasters gave identical high-quality sensory scores for all the samples evaluated, implying that the current process can be used to replicate factory conditions and achieve an acceptable good quality tea. Professional tea tasters are trained to classify different teas based on the overall quality, but not to identify

quality differences for different tea varieties. The present optimized miniature process can be used for manufacturing different tea varieties, however technologies that can help in achieving optimum wither and fermentation conditions for the many tea varieties under different environmental conditions could be explored further. Objective methods of evaluating fresh tea shoots polyphenols, caffeine, amino acids and aroma compounds could be considered and used in predicting high quality black tea for the newly developed varieties.

References

Wei, K., Wang, L., Zhou, J., He, W., Zeng, J., Jiang, Y., Cheng, H. (2011). Catechin contents in tea (Camellia sinensis) as affected by cultivar and environment and their relation to chlorophyll contents. *Food Chemistry, 125*(1), 44-48.

Mondal, T. K. (2014). *Breeding and biotechnology of tea and its wild species*: Springer Science & Business Media.

Tang, P., Shen, D.-Y., Xu, Y.-Q., Zhang, X.-C., Shi, J., Yin, J.-F. (2018). Effect of Fermentation Conditions and Plucking Standards of Tea Leaves on the Chemical Components and Sensory Quality of Fermented Juice. *Journal of Chemistry, 2018*, 1-7. doi:10.1155/2018/4312875

Zheng, X.-Q., Li, Q.-S., Xiang, L.-P., Liang, Y.-R. (2016). Recent advances in volatiles of teas. *Molecules, 21*(3), 338.

Ho, C.-T., Zheng, X., Li, S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9-27. doi:10.1016/j.fshw.2015.04.001

Payne, R. W., Murray, D. A., Harding, S. A., Baird, D. B., Soutar, D. M. (2009). GenStat for windows introduction. *VSN International, Hemel Hempstead, 204*.

Someswararao, C., Srivastav, P. P., Das, H. (2013). Quality of black teas in Indian market. *African Journal of Agricultural Research, 8*(5), 491-494.

Owuor, P. O., Obanda, M. (2007). The use of green tea (Camellia sinensis) leaf flavan-3-ol composition in predicting plain black tea quality potential. *Food Chemistry, 100*(3), 873-884. doi:10.1016/j.foodchem.2005.10.030

Scharbert, S., Holzmann, N., Hofmann, T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *Journal of Agricultural and Food Chemistry, 52*(11), 3498-3508.

Obanda, M., Owuor, P. O., Taylor, S. J. (1997). Flavanol composition and caffeine content of green leaf as quality potential indicators of Kenyan black teas. *Journal of the Science of Food and Agriculture, 74*(2), 209-215.

Hazarika, K. (2012). Domestic Consumer Market for Indian Tea: A Survey Report. *International Journal of Latest Trends in Finance and Economic Sciences, 2*(3), 251-256.

Bhuyan, L. P., Hussain, A., Tamuly, P., Gogoi, R. C., Bordoloi, P. K., Hazarika, M. (2009). Chemical characterisation of CTC black tea of northeast India: correlation of quality parameters with tea tasters' evaluation. *Journal of the Science of Food and Agriculture, 89*(9), 1498-1507. doi:10.1002/jsfa.3614

Chen, Q., Zhang, D., Pan, W., Ouyang, Q., Li, H., Urmila, K., Zhao, J. (2015). Recent developments of green analytical techniques in analysis of tea's quality and nutrition. *Trends in Food Science & Technology, 43*(1), 63-82. doi:10.1016/j.tifs.2015.01.009

Wang, J., Wang, Y., Cheng, J., Wang, J., Sun, X., Sun, S., Zhang, Z. (2018). Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy. *Lwt, 96*, 90-97. doi:10.1016/j.lwt.2018.05.012

Deb, S., Pou, K. R. J. (2016). A Review of Withering in the Processing of Black Tea. *Journal of Biosystems Engineering, 41*(4), 365-372.

Han, Z.-X., Rana, M. M., Liu, G.-F., Gao, M.-J., Li, D.-X., Wu, F.-G., Li, X.-B., Wan, X.-C., Wei, S. (2016). Green tea flavour determinants and their changes over manufacturing processes. *Food Chemistry, 212*, 739-748.

Owuor, P. O., Obanda, M. (2017). Clonal Variations in the Response of Hard Physically Withered Leaf to Rehydration Following Long Chemical Wither Durations.

Owuor, P. O., Obanda, M. (1997). Advances in withering technologies and future strategies to cope with high tea production in Kenya. *Tea, 18*(2), 184-193.

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 Chemistry, 75*(4), 395-404.

Owuor, P. O., Obanda, M. A., Othieno, C. O., Horita, H., Tsushida, T., Murai, T. (1987). Changes in the chemical composition and quality of black tea due to plucking standards. *Agricultural and Biological Chemistry, 51*(12), 3383-3384.

Liang, Y., Lu, J., Zhang, L., Wu, S., Wu, Y. (2003). Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions. *Food Chemistry, 80*(2), 283-290.

Wright, L. P., Mphangwe, N. I. K., Nyirenda, H. E., Apostolides, Z. (2002). Analysis of the theaflavin composition in black tea (Camellia sinensis) for predicting the quality of tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 82*(5), 517-525.

Obanda, M., Owuor, P. O., Mang'oka, R., Kavoi, M. M. (2004). Changes in thearubigin fractions and theaflavin levels due to variations in processing conditions and their influence on black tea liquor brightness and total colour. *Food Chemistry, 85*(2), 163-173.

Bhattacharyya, N., Bandyopadhyay, R., Bhuyan, M., Tudu, B., Ghosh, D., Jana, A. (2008). Electronic nose for black tea classification and correlation of measurements with "Tea Taster" marks. *IEEE transactions on instrumentation and measurement, 57*(7), 1313-1321.

Yu, P., Yeo, A. S., Low, M. Y., Zhou, W. (2014). Identifying key non-volatile compounds in ready-to-drink green tea and their impact on taste profile. *Food Chem, 155*, 9-16. doi:10.1016/j.foodchem.2014.01.046

Liang, Y. R., Ye, Q., Jin, J., Liang, H., Lu, J. L., Du, Y. Y., Dong, J. J. (2008). Chemical and instrumental assessment of green tea sensory preference. *International Journal of Food Properties, 11*(2), 258-272.

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

Zhang, L., Tai, Y., Wang, Y., Meng, Q., Yang, Y., Zhang, S., Yang, H., Zhang, Z., Li, D., Wan, X. (2017). The proposed biosynthesis of procyanidins by the comparative chemical analysis of five Camellia species using LC-MS. *Sci Rep, 7*, 46131. doi:10.1038/srep46131

Yashin, A. Y., Nemzer, B. V., Combet, E., Yashin, Y. I. (2015). Determination of the chemical composition of tea by chromatographic methods: a review. *Journal of Food Research, 4*(3), 56.

Fraser, K., Lane, G. A., Otter, D. E., Harrison, S. J., Quek, S. Y., Hemar, Y., Rasmussen, S. (2013). Monitoring tea fermentation/manufacturing by direct analysis in real time (DART) mass spectrometry. *Food Chem, 141*(3), 2060-2065. doi:10.1016/j.foodchem.2013.05.054

Mishra, P., Nordon, A., Tschannerl, J., Lian, G., Redfern, S., Marshall, S. (2018). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering*.

Yaroshenko, I., Kirsanov, D., Kartsova, L., Bhattacharyya, N., Sarkar, S., Legin, A. (2014). On the application of simple matrix methods for electronic tongue data processing: Case study with black tea samples. *Sensors and actuators B: Chemical, 191*, 67-74.

Sharma, P., Ghosh, A., Tudu, B., Sabhapondit, S., Baruah, B. D., Tamuly, P., Bhattacharyya, N., Bandyopadhyay, R. (2015). Monitoring the fermentation process of black tea using QCM sensor based electronic nose. *Sensors and actuators B: Chemical, 219*, 146-157.

Roy, R. B., Tudu, B., Shaw, L., Jana, A., Bhattacharyya, N., Bandyopadhyay, R. (2012). Instrumental testing of tea by combining the responses of electronic nose and tongue. *Journal of Food Engineering, 110*(3), 356-363.

Gill, G. S., Kumar, A., Agarwal, R. (2011). Monitoring and grading of tea by computer vision–A review. *Journal of Food Engineering, 106*(1), 13-19.

Ning, P., Gong, C., Zhang, Y., Guo, K. (2010). La, Ce, Pr, Nd and Sm concentrations in Pu'er tea of Yunnan, China. *Journal of Rare Earths(English Edition), 28*(4), 636-640.

Zhu, H., Ye, Y., He, H., Dong, C. (2017). Evaluation of green tea sensory quality via process characteristics and image information. *Food and Bioproducts Processing, 102*, 116-122. doi:10.1016/j.fbp.2016.12.004

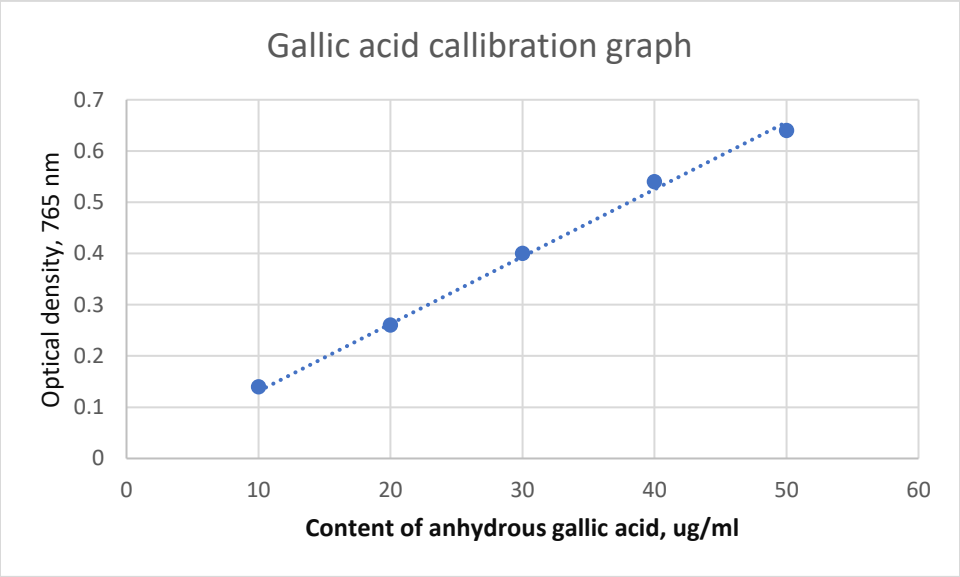

Figure S5. 1: Gallic acid calibration graph.

Determinants of tea quality – multivariate analyses of biochemical compounds in fresh leaf and organoleptic properties

Lubanga M. Nelson[b,c*], De-Silva Jacquie[d], Massawe Festo[b], Mayes Sean[a]

*nelsonlubanga@gmail.com

[a]School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK.

[b]School of biosciences, University of Nottingham, Malaysia Campus Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[c]Unilever Tea Kenya Ltd., Nakuru Road, Kericho, 20-20200, Kenya.

[d]Unilever Research & Development, Colworth Science Park, Sharnbrook, Bedford MK441LQ, UK.

Abstract

Tea (*Camellia sinensis* (L) O. Kuntze) quality is influenced by the levels of polyphenols, aroma compounds, caffeine and amino acids present in fresh tea shoots. These compounds are also influenced by genotype, environment, post-harvest handling and climate. The present study was conducted to evaluate and select high quality tea varieties using biochemical data. We also correlated biochemical data with sensory evaluation data. Significant differences ($P<0.01$) among genotypes were observed for all the traits. However, G x E was not significant ($p<0.05$) for all the traits. Multivariate analysis using hierarchical cluster and principle component analyses successfully differentiated the clones based on their biochemical properties. For example, clones Clone97, Clone70 and Clone53 showed high levels of epicatechin gallate (ECG), theogallin, gallic acid (GA) and gallocatechin gallate (GCG) across the sites. Sensory evaluation revealed Clone69, Clone97 and Clone70 as high 'mouthfeel' varieties, while high 'taste' clones were Clone69 and Clone11. Mouthfeel and 'colour' were positively correlated ($r =0.66$). EGCG (Epigallocatechin gallate), EGC (Epigallocatechin gallate and EC (Epicatechin) were positively correlated with each other, and clustered together on the heat map, matching with sensory evaluation data on high 'taste'. By contrast, ECG was grouped with theogallin, C (catechin), GC (gallic acid), GCG and GA,

corresponding with high 'mouthfeel' scores. Results of this study are very promising and suggest that it could be possible to select high quality teas using only biochemical data in a tea breeding programme, hence saving time of field planting and miniature manufacture.

Keywords: Tea quality; multivariate analysis, biochemical compounds; sensory evaluation

6.1. Introduction

Tea (*Camellia sinensis* (L) O. Kuntze) is the most common non-alcoholic beverage in the world after water (Cheng, 2004). It is also a major source of livelihood in many tropical and sub-tropical countries, as it has led to the creation of jobs for many people and development of good social infrastructure such as road networks, schools and hospitals in rural areas (Mukhtar et al., 2000b). For instance, Kenya is the world's leading black tea exporter (ITC, 2017), contributing the highest foreign exchange earnings to its' economy (TBK, 2017). The tea sector contributed about 4% of Kenya's GDP and 26% of the country's total export earnings in 2017 (TBK, 2017).

Quality is among the most important attributes in a tea breeding programme besides yield, drought tolerance and resistance to pest and diseases (Corley et al., 2018; Mondal, 2014b; Kamunya et al., 2012; Jain et al., 2009). It is the main basis for pricing at the tea auction (Gallaher et al., 2006), and is characterized by flavour and colour (Zheng et al., 2016a; Ho et al., 2015; Yang et al., 2013; Kumar et al., 2011b; Wang et al., 2010). Flavour (taste, mouthfeel and aroma) and colour (liquor and appearance of dry made tea) are usually assessed during sensory quality evaluation (Ho et al., 2015; Chaturvedula et al., 2011; Wang et al., 2010). Taste, mouthfeel and colour are generated by non-volatile compounds (Chen et al., 2008c; Kawakami, 1997; Harbowy et al., 1997; Yamanishi, 1995), while aroma is produced by volatile compounds in tea (Zheng et al., 2016a; Ho et al., 2015; Yang et al., 2013).

Green leaf polyphenols, aroma compounds, caffeine and amino acids are the main compounds known to influence the quality of black tea (Tu et al., 2018; Koch et al., 2018; Magagna et al., 2017; Jeganathan et al., 2017; Zheng et al., 2016a; Ho et al., 2015). The four main catechins in fresh and young green tea leaf are (-)-epigallocatechin-3-gallate (EGCG), (-)-epigallocatechin (EGC), (-)-epicatechin-3-

gallate (ECG) and epicatechin (EC) (El-Shahawi et al., 2012b; Liang et al., 2003; Nakabayashi, 1994), while the minor catechins include (+)-Gallo catechin (GC), (-)-Gallo catechin gallate (GCG), (-)-catechin gallate (CG) and (+)-catechin (C) (Chen et al., 2010; Friedman et al., 2005; Yamamoto et al., 1997). Gallated catechins include EGCG, ECG and GCG while non-gallated catechins are EC, C, EGC and GC (Fan et al., 2016). Biochemical compounds in tea are routinely analysed using analytical techniques such as high-performance liquid chromatography, capillary electrophoresis, mass spectrometry and nuclear magnetic resonance (Samynathan et al., 2016; Lee et al., 2014a; Lee et al., 2011). Analytical techniques are believed to be more reliable for determining tea quality when compared to the conventional sensory evaluation method that is time time-consuming, laborious (Bhattacharyya et al., 2008) and highly subjective (Yu et al., 2014). In this study, NMR was preferred over other analytical techniques because it can simultaneously quantify levels of catechins, alkaloids and amino acids and is accurate, reproducible, non-destructive and requires minimal sample preparation steps (Kumar et al., 2015; Yuan et al., 2014). ^{1}H NMR spectroscopy has been used previously to study the effects of climatic conditions on green tea metabolites in three different growing areas (Lee et al., 2010) and to simultaneously quantify catechins, caffeine, theanine and gallic acid (Yuan et al., 2014). The objective of this study was to evaluate the quality of tea genotypes using biochemical data obtained from NMR and determine the correlation between the measured green leaf biochemical compounds and organoleptic evaluation results.

6.2. Materials and methods

6.2.1. Plant materials

In the present study, 103 and 106 genotypes (Table S6. 3 and Table S6. 4) were selected in clonal field trials (CFTs) present at Jamji and Sites respectively. Several known high quality standard clones were included such as Clone97 and CLONE105 (Kamunya et al., 2012) alongside CLONE101, CLONE95, Clone61, 10/3, Clone62, Clone63 and Clone60 (Gabriel Tuwei, Pers. com, June 2016, R&D UTK). A complete description of the quality standard checks and commercial varieties used in this study is shown in Table 7. The crosses for all the varieties used in this study is presented in Table S6. 1. Green leaf samples

at each location were collected at three different times of the year; in April 2016, December 2016 to March 2017 and May to June 2017.

6.2.2. Test locations and experimental designs

The study was conducted in two locations within UTK: at Kericho estate and at Jamji estate. Kericho estate is located at 0^0 22' S and 35^0 17' E, 2005 meters above sea level (m.a.s.l). Jamji estate is located at 0^0 28' S and 35^0 11'E, 1733 m.a.s.l. The descriptions of the climate are presented in Figure 7 and 8. The mean average temperature and rainfall range were 18.8^0c and 1735 mm, at Kericho and 18.01^0c and 1935 mm respectively, at Jamji. However, there are variations between monthly temperatures throughout the year (Figures 7 and 8).

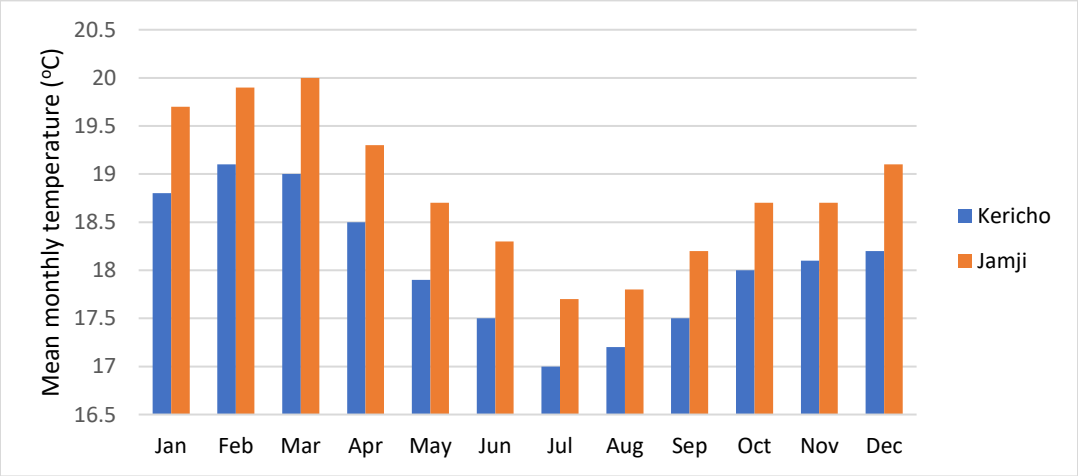

Figure 7. Mean temperatures at Kericho and Jamji.

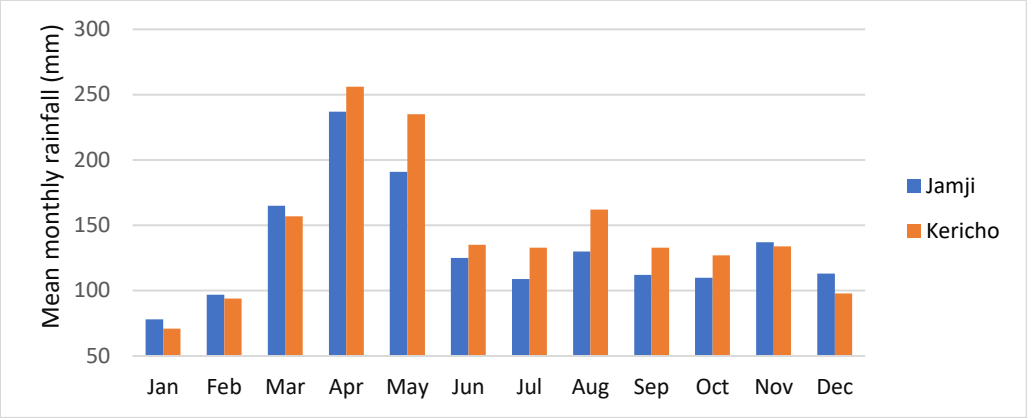

Figure 8. Mean monthly rainfall at Kericho and Jamji.

6.2.3. Leaf Sampling and sample processing

About 200g of fresh leaf (comprising of the youngest two leaves and a bud) were plucked from each of the clones in the CFTs and placed in the cooler box before transporting to the laboratory. The samples were placed in the deep freezer at -20^{0}C for 48 hours and then put in the freeze drier for a further 48 hours, to dry and preserve the samples, and prevent oxidation. The freeze-dried samples were sent to Rothamsted research laboratories in United Kingdom (UK) for biochemical analysis using NMR.

6.2.4. NMR methodology

Determination of the quantities of theobromine, caffeine, gallic acid (GA), epicatechin (EC), gallocatechin gallate (GCG), epicatechin gallate (ECG), epigallocatechin (EGC), epigallocatechin gallate (EGCG), catechin (C), gallocatechin (GC) and theanine was carried out by the following method described by Yuan et al. (2014) at Rothamsted research laboratories in United Kingdom (UK). The freeze-dried leaf was crashed into a fine powder using a coffee grinder. Fifteen grams of each of the samples was weighed carefully and mixed with an extraction solvent (80:20 D2O:CD3OD containing 0.01% w/v DSS, 1mL). The mixture was heated at 500C for 10 minutes, allowed to cool and centrifuged at 10,000 rpm for 10 minutes. The supernatant was removed, and the remaining mixture heated at 90^{o}C for 2 minutes. The samples were cooled to 40^{0}C for 30 minutes and centrifuged. 700uL was removed and put in a clean tube.

Potassium phosphate buffer (pH 5.8, 2.4M, 40uL) was added, mixed with EDTA (32mM, 20uL) and centrifuged. 650 uL of the supernatant was removed, placed in a 5mm NMR tube and analysed using 1H NMR. The mean NMR data across the sites, Jamji and Kericho is presented in Supplementary Table S6. 2, Table S6. 3, and Table S6. 4, respectively.

6.2.5. Sensory evaluation

Eight genotypes and their 4 biological replicates were evaluated in this study. The 8 varieties comprised of 2 contrasting groups that were selected based on their extreme total catechin levels. The varieties with high levels of total catechins include Clone97, CLONE101, Clone69 and Clone70, while those with low levels of total catechins are Clone11, Clone15, CLONE104 and CLONE96. Each of the 8 clones was harvested, every day for 4 days between 8am – 9am. Variabilities were minimised by plucking the youngest two leaves and an apical bud, while handling and transporting in the same way and time prior to processing. The 4 biological replicates of each of the 8 clones were then manufactured into black tea separately and subjected to sensory evaluation as follows; 5.6 grams of the sample was measured using a hand scale and transferred into an infusion pot. De-ionized water was boiled and added into the infusion pot until it was full. The boiled water and sample were left to brew for 6 minutes. Using a syringe, 5 ml of milk was carefully placed into each tasting bowl. The infused liquor was then filtered into the tasting bowl containing milk and left to stand for 10 minutes before it was analysed. The residue from the infusion was collected on the infusion cup lid. The liquor was then analysed accordingly. Quality of the samples was blindly assessed by a trained, experienced and calibrated professional taster at Unilever R&D Colworth in United Kingdom. The taster evaluated the samples based on taste, mouthfeel and colour, which are the main desirable quality traits in tea.

6.2.6. Data and statistical analysis

Analysis of variance across the sites and at each individual site was carried out using GenStat statistical software, 18th Edition (Payne, 2015), using the following statistical model for the combined data across the sites:

$$Yijkl = \mu + Gi + Ej + GEij + Rk(j) + Bl(k) + \varepsilon ijkl$$

Where: Yijkl is the measured value of the clone i in block l and replication k of environment j, μ is the mean, Gi is the effect of genotype i, Ej is environment effect, GEij is the genotype by environment interaction i with environment j, Rk(j) is the replication effect k in environment j, Bl(k) is the effect of block l in replication k, ϵijkl is the effects of residual of genotype i in block l and replication k of environment j.

Significant differences among sample means were compared using Duncan's multiple range test at $p<0.05$ level of significance. Data was expressed as means ± SD of the 3 biological replicates with $P < 0.05$ representing a statistically significant differences between the main treatments. Correlations among the phenotypic traits and between sensory data were done using excel. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) was performed at each site and across the two sites for the biochemical data using Qlucore omics Explorer 3.4. A heatmap showing two-dimensional clustering of variables and clonal set was produced, where the degree of cluster associations in the form of dendrograms were indicated at the top (clonal grouping) and on the left (grouping of phenotypes). A clone by 12-dimensional matrix of the biochemical data set was displayed on the heatmap where each column represents the standardized binary measurements for a given clone at Kericho, Jamji and across the sites. High values for the traits are coded red while low values were coded as green.

Table 7. A description of the high-quality standard checks and commercially planted varieties in Eastern and Southern Africa that were used in this study.

	Variety	History/parentage	Status	Properties	Source
1	CLONE105	Field selection obtained from Chomogonday estate, Kericho	Commercial	High quality	(Kamunya et al., 2012)
2	CLONE106	Open pollinated cross of TRFK 91/1 (camellia irrawadiensis)	Commercial	High anthocyanin rich tea	(Kamunya et al., 2012)
3	CLONE102	Open pollinated progeny of AHP CLONE100	Commercial	High yield and quality, tolerant to mites and nematodes	(Kamunya et al., 2012)
4	CLONE103	TRFCA SFS150 x EPK TN14-3	Commercial	High and quality	(Kamunya et al., 2012)
5	TRFK Clone67	Field selection obtained from Kimugu estate, Kericho	Commercial	Medium quality	(Kamunya et al., 2012)
6	TRFK Clone66	open pollinated variety of CLONE105	Commercial	Medium quality	(Kamunya et al., 2012)
7	TRFK Clone65	open pollinated variety of CLONE105	Commercial	Medium quality	(Kamunya et al., 2012)
8	Clone97	Field selection obtained from George Williamson	Commercial	High quality	(Kamunya et al., 2012)
9	10/3	UTK field selection	Experimental	High quality	UTK
10	Clone63	UTK field selection	Experimental	High quality	UTK
11	Clone61	UTK field selection	Experimental	High quality	UTK
12	Clone62	UTK field selection	Experimental	High quality	UTK
13	Clone60	UTK field selection	Experimental	High quality	UTK
14	CLONE95	UTK field selection	Commercial	High quality	UTK
15	CLONE101	Originated from James Finlays of Kenya	Commercial	High quality	UTK
16	TRFCA MFS 87	Malawian assam selection	Commercial	High yielding	(Kamunya et al., 2012)
17	TRFCA SFS 150	Malawian assam field selection	Commercial	Drought tolerance, high yield and quality	(Kamunya et al., 2012)
18	TRFCA PC 81	Malawian assam selection	Commercial	Moderate quality	(Kamunya et al., 2012)
19	TRFCA PC 108	Malawian assam selection	Commercial	Moderate quality	(Kamunya et al., 2012)
20	UTK 95/765	PC110 x CLONE98	Commercial	High yielding s	UTK
21	CLONE8	CLONE98 x AHP CLONE100	Commercial	High yielding	UTK
22	CLONE37	TRFK Clone65 x AHP CLONE100	Commercial	High yielding	UTK
23	UTK 00/2444B	SFS150 x Clone105	Commercial	High yielding recommended in high altitudes	UTK
24	UTK 00/4803B2	PC110 x CLONE98	Commercial	High yielding	UTK
25	CLONE98	Seedling selection	Commercial	Drought tolerant, moderate yielding	UTK
26	CLONE96	Seedling selection of China origin	Experimental	Poor black tea quality	UTK

6.3. Results

6.3.1. Biochemical evaluation at Kericho and Jamji

The mean biochemical data for each of the 12 traits at Jamji and Kericho is presented in Figure 9. Out of the 12 traits, 9 traits recorded higher mean biochemical levels at Jamji compared to Kericho. However, the difference was not large as can be seen from the error bars (Figure 9). Analysis of variance table and means of biochemical traits at Kericho and Jamji are presented in Tables 8 and 9, respectively. Significant differences ($P<0.01$) were observed between the genotypes for all traits except Epicatechin (EC) at both locations (Tables 8 and 9).

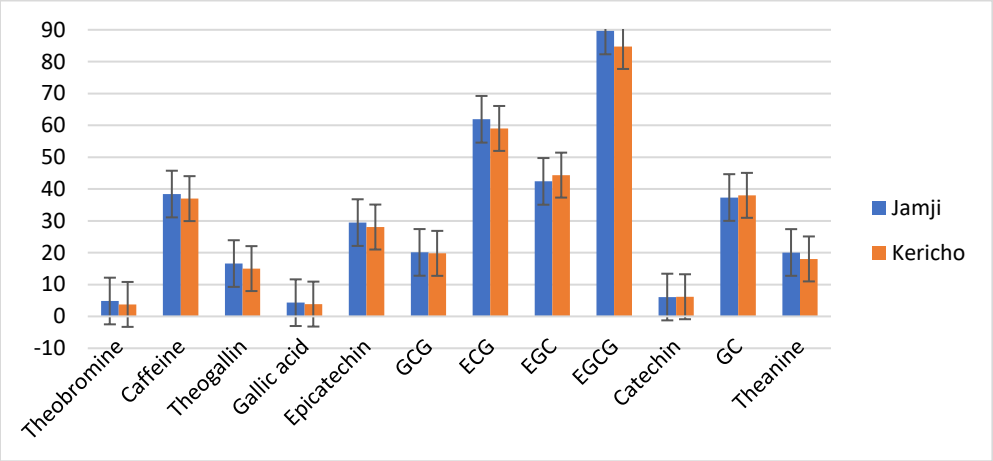

Figure 9. Comparison of the mean biochemical traits at Kericho and Jamji.

A positive and significant ($p<0.05$) correlation of 0.99 was observed between the mean biochemical compounds across the two sites (Figure 10). Correlation coefficients (r) among the traits at Kericho and Jamji are displayed on Tables 10 and 11, respectively. At both sites, positive and significant correlations were observed between GCG and GA, ECG and theogallin, EGC and EC, EGCG and EC, catechin and ECG, ECG and GC, and GC and catechin (Tables 10 and 11). Similarly, significant ($p<0.05$) negative correlations at both sites were observed between theogallin and EGC, EGC and GC, ECG and EGC, and EGC and catechin (Table 10 and 11). Combined analysis of variance showed that genotype was highly

significant (P<0.01) for all the traits (Table 12). Environment (E) was significant (p<0.05) for all traits except GC and GCG, while all traits were not significantly influenced by G x E (p<0.05) (Table 12).

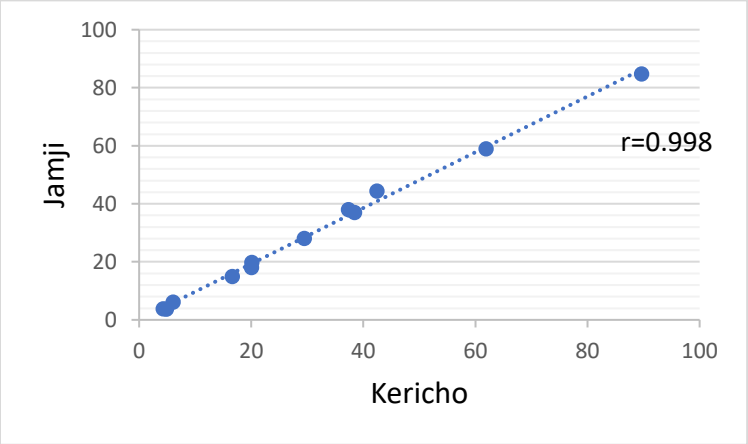

Figure 10. Correlations of the mean biochemical traits between Jamji and Kericho.

Table 8. Analysis of variance and means of the biochemical traits at Kericho

Source	Theobromine	Caffeine	Theogallin	GA	Epicatechin	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Genotype	3.58***	41.67***	19.37***	1.67***	33.52 ns	52.79***	530.56***	449.19***	470.1***	18.97***	351.78***	64.77***
CV (%)	36.1	9.1	18.2	20.6	19.5	16.4	11.3	17.2	14.2	49.1	16.9	30.2
LSD (p<0.05)	2.1891	5.412	4.398	1.2826	8.823	5.225	10.724	12.274	19.408	4.884	10.349	8.762
Mean	3.77	36.99	15.02	3.87	28.06	19.80	59.03	44.37	84.77	6.18	38.02	18.05
Maximum	6.25	46.25	24.09	7.73	37	44.25	113.05	65.36	128.6	14.38	70.13	29.51
Minimum	1.53	27	10.55	2.73	19.75	12.66	39.96	14.28	61.09	2.34	20.58	8.42

Table 9. Analysis of variance and means of the biochemical traits at Jamji

Source	Theobromine	Caffeine	Theogallin	Gallic acid	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Genotype	7.65***	64.76***	19.85***	1.36**	29.96 ns	29.51***	541.82***	439.9***	523***	21.39**	367.46***	67.18***
CV (%)	37.5	10.6	18.6	21.9	20	19.2	15.3	20.1	15.3	61.90	19.7	23.6
LSD (p<0.05)	2.922	6.58	4.976	1.5195	9.482	6.2	15.287	13.7	22.116	6.06	11.834	7.612
Mean	4.84	38.42	16.60	4.31	29.46	20.10	61.91	42.40	89.64	6.08	37.33	20.08
Max	10.88	47.06	24.70	6.66	37.27	28.77	112.16	73.97	115.85	15.53	69.01	32.13
Minimum	1.62	16.01	11.14	2.92	20.09	12.54	36.89	18.18	55.66	2.83	9.40	9.59

Table 10. Correlations between the biochemical traits at Kericho

Traits	Theobromine	Caffeine	Theogallin	Gallic acid	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Theobromine	1											
Caffeine	0.47	1										
Theogallin	0.24	0.10	1									
Gallic acid	0.00	-0.02	0.43	1								
Epicatechin	0.27	0.31	0.17	0.39	1							
GCG	-0.03	-0.05	0.44	0.96***	0.35	1						
ECG	0.08	0.04	0.84***	0.31	-0.12	0.37	1					
EGC	-0.11	-0.07	-0.50*	0.33	0.48	0.30	-0.60*	1				
EGCG	0.33	0.38	0.02	0.00	0.77***	-0.08	-0.24	0.35	1			
Catechin	0.07	0.02	0.32	-0.21	-0.42	-0.17	0.57*	-0.55	-0.36	1		
GC	0.05	0.02	0.32	-0.26	-0.48	-0.21	0.67**	-0.63	-0.34	0.80***	1	
Theanine	0.25	0.22	-0.04	-0.13	-0.15	-0.09	-0.05	-0.12	-0.19	0.04	-0.05	1

Table 11. Correlations between biochemical traits at Jamji

Traits	Theobromine	Caffeine	Theogallin	Gallic acid	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Theobromine	1											
Caffeine	0.23	1										
Theogallin	0.29	0.10	1									
Gallic acid	0.26	-0.12	0.34	1								
Epicatechin	0.26	0.40	0.22	0.39	1							
GCG	0.07	-0.16	0.27	0.86	0.25	1						
ECG	0.02	0.12	0.75***	0.14	-0.09	0.25	1					
EGC	-0.06	-0.07	-0.46	0.33	0.51*	0.39	-0.60**	1				
EGCG	0.26	0.44	0.04	0.13	0.84***	-0.08	-0.26	0.42	1			
Catechin	0.00	0.13	0.32	-0.22	-0.31	-0.11	0.64**	-0.52*	-0.31	1		
GC	-0.14	0.24	0.28	-0.36	-0.31	-0.22	0.71***	-0.58	-0.36	0.71***	1	
Theanine	0.04	0.14	-0.06	-0.27	-0.14	-0.26	-0.05	-0.10	-0.07	0.14	0.12	1

*, **, *** correspond to significant, moderate significant and high significant correlations.

Table 12. Mean squares from the combined ANOVA for biochemical compounds across the two sites

Source of variation	Theobromine	Caffeine	Theogallin	Gallic acid	Epicatechin	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Genotype	8.21***	77.82***	33.02***	2.1648***	44.59**	66.51***	977.99***	809.26***	844.8***	30.2***	616.04***	94.18***
Environment (E)	160.77***	433.98***	406.83***	28.65***	341.05***	18.3ns	1623.19***	558.94**	3936.8***	0.42*	13.25 ns	666.04***
G x E	2.1 ns	13.93 ns	5.54 ns	0.69 ns	16.83 ns	14.44 ns	86.29 ns	63.27 ns	136.7 ns	10.58 ns	88.41ns	38.09 ns
Error	2.74	14.48	8.56	0.86	31.87	13.06	68.83	67.26	163.8	12.09	48.67	32.59
CV (%)	38.7	10.1	18.5	22.8	19.6	18.1	13.7	18.9	14.7	56.5	18.4	29.9
LSD (p<0.05)	2.66	6.107	4.7	1.49	9.06	5.8	13.316	13.16	20.54	5.58	11.20	9.163
Variation explained (%)	61.3	70.47	81.5	77.2	80.4	81.9	83.63	81.1	85.3	45.85	81.6	70.1
Mean	4.28	37.83	15.83	4.08	28.81	19.97	60.65	43.42	87.29	6.15	37.87	19.09
maximum	6.84	46.66	24.39	5.82	34.95	33.01	112.61	69.3	120.82	12.89	63.28	29.82
Minimum	2.1	27.92	11.47	2.94	22.51	13.65	40.51	17.13	59.89	2.93	22.36	10.7

* Significant at P < 0.05.
** Significant at P < 0.01.
*** Significant at P < 0.001.
ns, not significant, GCG, gallocatechin gallate, ECG, epicatechin gallate, EGC, epigallocatechin, EGCG. Epigallocatechin gallate, GC, gallocatechin.

6.3.2. Multivariate analysis

6.3.2.1. PCA plot for the biochemical data at Kericho

A PCA plot was generated with samples coloured by clone. The first, second and third principal components (PCs) accounted for 32%, 24% and 17% of the total variance respectively. The score plot of the first three principle components showed eight clusters that could be distinguished from one another (Figure 11).

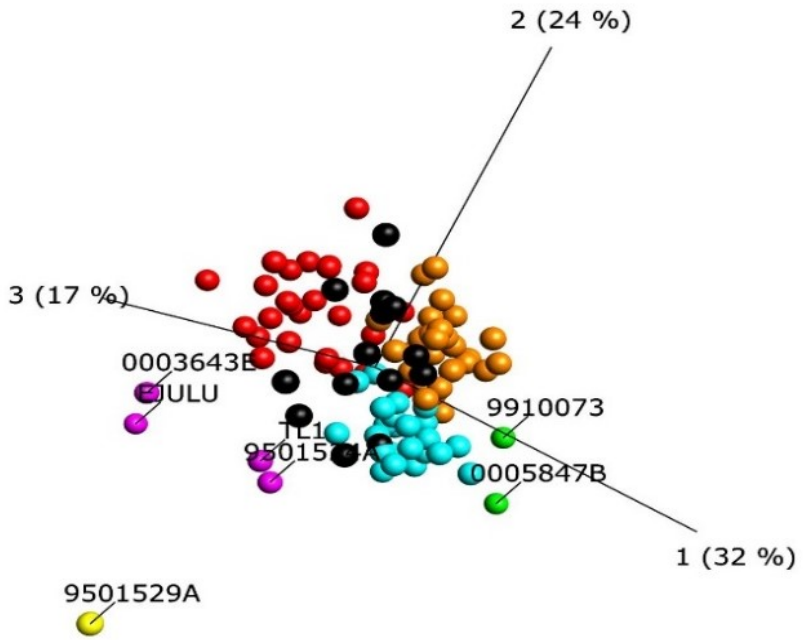

Figure 11. A PCA plot generated with samples coloured by different groups for the 103 genotypes at Kericho.

6.3.2.2. Hierarchical cluster analysis of the biochemical traits at Kericho site

Hierarchical cluster analysis differentiated the varieties into 8 major groups (Figure 12). The first cluster comprised of only Clone70 and it had high levels of ECG, theogallin, GCG and GA (Figure 12). The second cluster included clones Clone97, CLONE101, Clone69 and Clone53 (Figure 12), and had high levels of GCG, gallic acid, ECG, theogallin, GC, catechin, and low levels of EGC, EGCG, EC, caffeine and theobromine (Figure 12). The seventh cluster was comprised of 24 varieties, among them, four known

quality standard checks Clone62, CLONE95, Clone60 and Clone63, and had high levels of EGCG and EC (Figure 12).

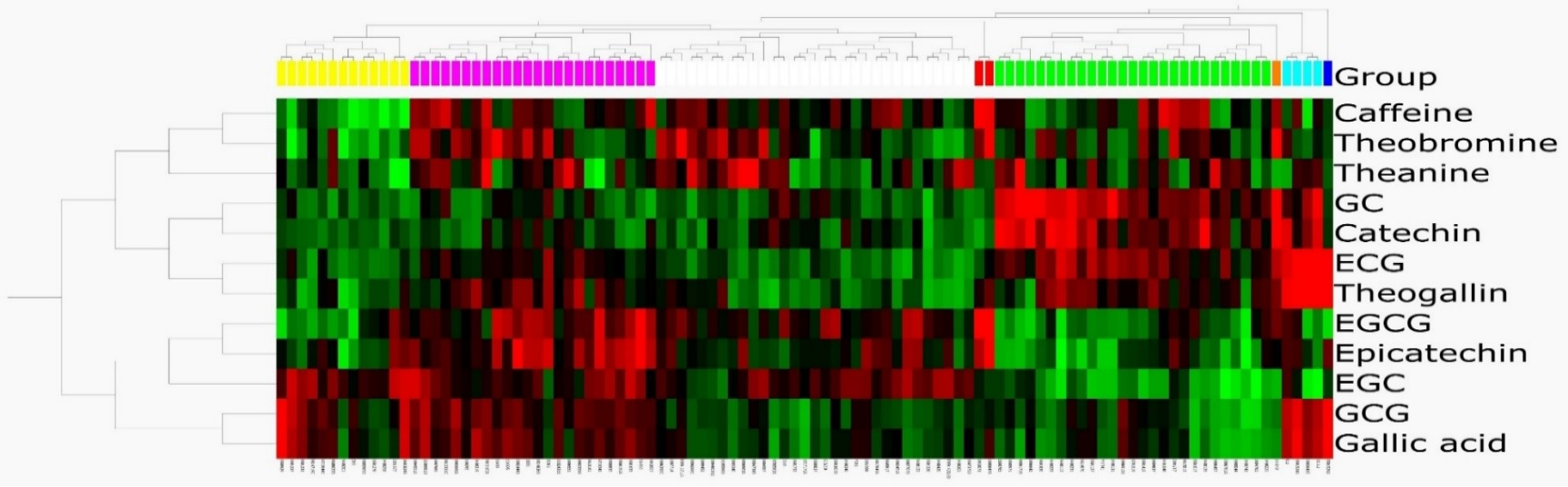

Figure 12. HCA dendrogram of the 103 varieties and all the measured traits at Kericho.

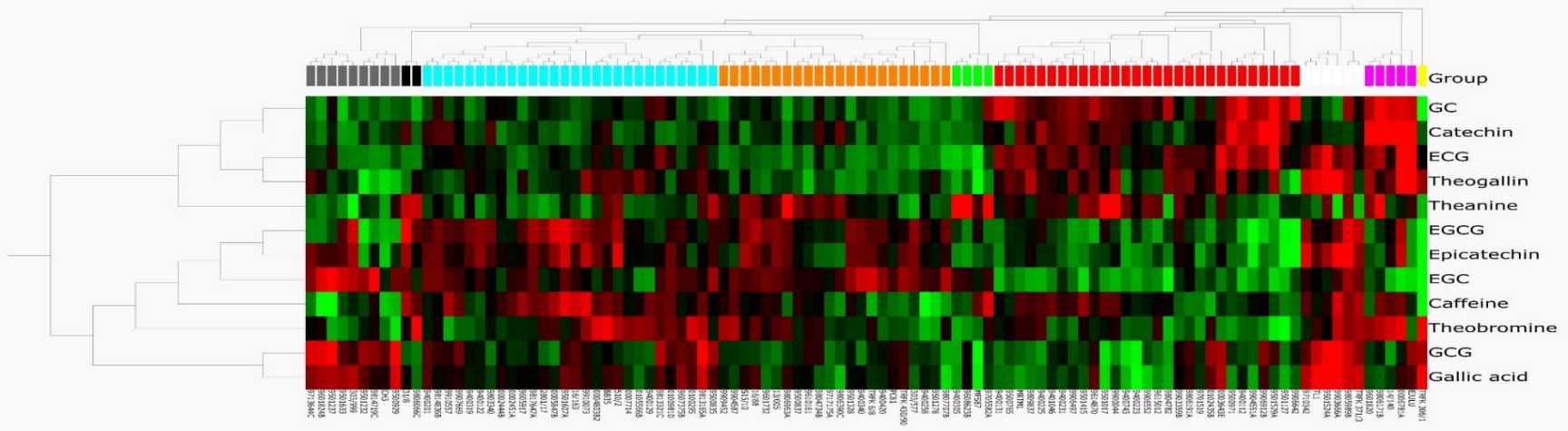

Figure 13. HCA dendrogram of the 106 varieties and the 12 measured traits at Jamji.

6.3.2.3. Principle component analysis (PCA) of the biochemical traits at Jamji

A PCA plot was generated with samples coloured by clone and it explained 70% of the existing genetic diversity between three axes (Figure 14).

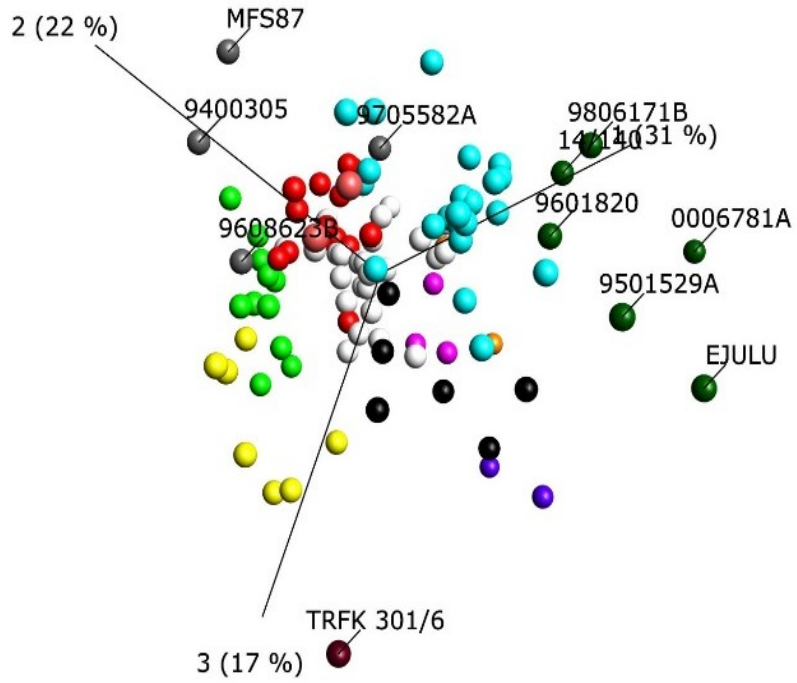

Figure 14. Principle component analysis (PCA) of 106 tea varieties based on the 12 biochemical traits at Jamji.

6.3.2.4. Hierarchical cluster analysis of the biochemical traits at Jamji

Hierarchical cluster analysis differentiated the genotypes into 9 main groups (Figure 13) and this was similar to the results of the PCA (Figure 13). The first group comprised of only clone CLONE106, a purple coloured clone with high levels of GA and theobromine, but low levels of caffeine, EGC, EC, EGCG and GC (Figure 13). The second group comprised of Clone97, Clone81, Clone61, Clone31, Clone56, and had high levels of GC, catechin, ECG, theogallin and theobromine. The third cluster comprised of clone Clone92, Clone69, Clone79, CLONE101, CLONE37 and CLONE102 (Figure 13), and had high levels of theogallin, EC, GCG and GA. The seventh cluster had 28 clones that had high levels of EGCG, EC, caffeine and theobromine, but low levels of GC and catechins.

6.3.2.5. Principle component analysis (PCA) of the biochemical traits across the two sites

Principle component analysis differentiated the varieties into 8 major groups (Figure 15). A PCA plot was generated from all the traits investigated and it explained 75% of the existing variation, suggesting considerable diversity among the genotypes.

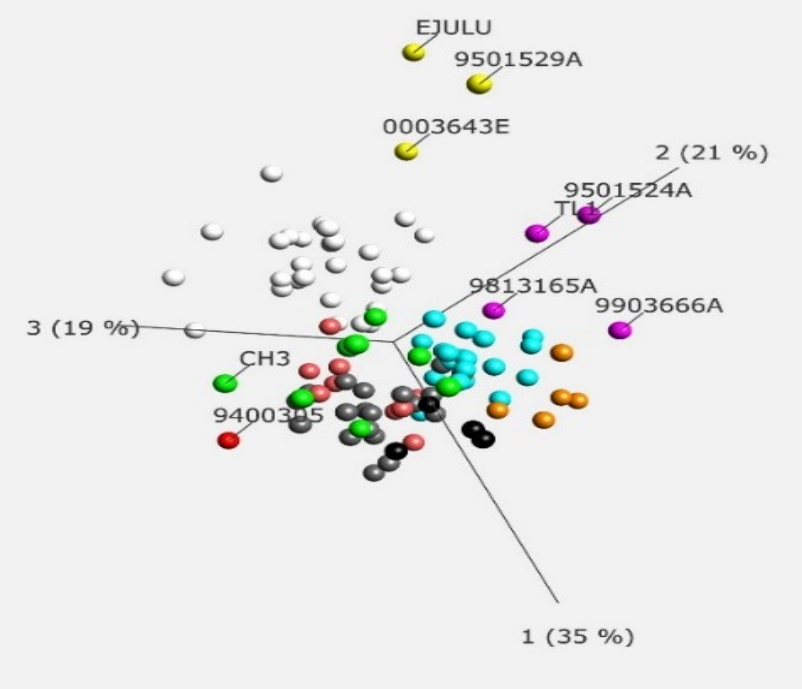

Figure 15. PCA of the 12 biochemical compounds for all the varieties at Kericho and Jamji.

6.3.2.6. Hierarchical cluster analysis across the sites

Hierarchical cluster analysis for the mean biochemical data separated the varieties into 8 major groups. The first group consisted of Clone97, Clone70 and Clone53. This group had high gallic acid, GCG, catechin, GC, ECG and theogallin, and low levels of EC, EGCG, EGC, caffeine and theobromine (Figure 16). The second cluster comprised of CLONE101, Clone69, Clone86 and Clone92, and it had high GA, GCG, ECG and theogallin (Figure 16). The sixth cluster comprised of 27 varieties, and had high EGCG, EC, caffeine and theobromine (Figure 16).

.

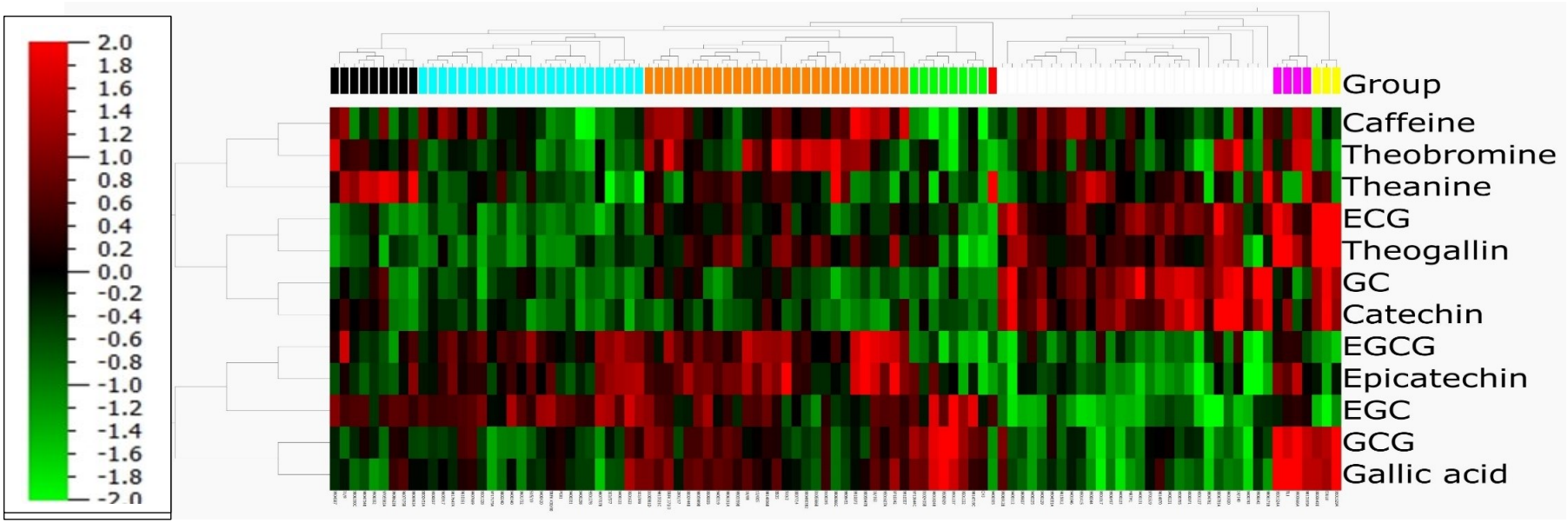

Figure 16. Heatmap showing the standardized 103 by 12-dimensional matrix of the biochemical data set across the two sites

Table 13. Correlations of the organoleptic evaluation results and biochemical traits

Traits	TB	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine	Colour	Taste	Mouthfeel
Colour	0.11	-0.61	0.41	0.46	-0.1	0.48	0.48	-0.26	-0.39	0.51*	0.28	-0.5*	1		
Taste	0.06	0.15	-0.35	-0.2	0.32	-0.29	-0.51	0.81***	0.61	-0.78***	-0.83***	0.5*	-0.39	1	
Mouthfeel	0.09	-0.29	0.8	0.87***	0.47	0.87***	0.78	-0.09	-0.05	0.48	0.32	-0.27	0.66*	-0.09	1

6.3.3. Organoleptic tea quality evaluation

The taster considered all the samples acceptable and made no general adverse comments about the manufacture of the samples. There were significant ($p<0.01$) differences between the clones for taste and mouthfeel (Table 14). However, no significant ($p<0.05$) clonal differences were observed for colour (Table 14). Clones Clone69 and Clone97 were scored as high mouthfeel teas, while CLONE96 and Clone97 had high colour scores (Table 14). Clones that recorded high taste scores include Clone69, Clone11 and CLONE104 (Table 14). Correlations between the organoleptic results and the biochemical data was conducted. Mouthfeel had a strong and positive correlation with theogallin (0.80), GA (0.87), ECG (0.78) and GCG (0.87) (Table 13). Taste had a strong and positive correlation with EGC (0.81) and EGCG (0.61). Moderate correlations were also observed between colour and GCG (0.48), ECG (0.48) and catechin (0.51) (Table 13). Taste was negatively correlated with ECG (-0.51), catechin (-0.78) and GC (-0.83) (Table 13).

Table 14. Mean sensory evaluation results

Clone	Female	Male	Colour	Taste	Mouthfeel
Clone11	SFS150	CLONE100	4.5	5.3	4.6
Clone15	CLONE104	CLONE98	4.5	4.7	4.3
Clone69	CLONE101	Clone97	4.6	5.3	4.9
Clone70	PC110	Clone97	4.6	4.5	4.8
CLONE96	-	-	4.9	4.9	4.7
Clone97	-	-	4.9	4.5	4.9
CLONE104	-	-	4.4	5.0	4.3
CLONE101	PMC59	CLONE100	4.7	4.9	4.7
p values			0.3164	0.007	0.0095

6.4. Discussion

In this study, NMR spectroscopy was used to quantify biochemical compounds in tea. Variations among the clones for all the biochemical traits suggest the presence of genetic diversity that can be exploited in breeding. For each trait, several clones had significantly higher biochemical content than others and could be considered further for breeding of high-quality teas. The biochemical composition of fresh tea shoots

depends on the type of tea variety, climatic conditions, soils, altitude of the planting location, agronomic practices, harvesting/plucking season, quality of harvested fresh shoots, processing parameters and storage conditions (Yuan et al., 2014; Lee et al., 2010).

To inform the selection of high-quality genotypes, correlations between the biochemical traits and the sensory evaluation results was conducted to determine the most important biochemical traits linked to sensory tea quality. Earlier studies have shown that catechins, caffeine and amino acids influence the organoleptic properties of tea (Scharbert et al., 2005; Liang et al., 2003; Mahindroo, 2000). The biochemical contents of the known high-quality standard varieties and their grouping on the PCA plot and hierarchical cluster analysis were also compared with the sensory evaluation results. Selection of high-quality tea genotypes was done by analysing the main compounds influencing tea quality.

Tea quality is an important attribute as it significantly influences the price of tea at the tea auction. The taste of tea is usually recorded at the tea auction by the astringency and bitterness (Narukawa et al., 2011; Narukawa et al., 2010; Lee et al., 2007; Scharbert et al., 2004), while mouthfeel is described as the thickness of the infused tea. According to Chen et al. (2010), there is a strong correlation between the intensity of taste of tea and the catechins concentration. The gallated catechins (EGCG and ECG) are more astringent and give tea the distinctive bitter taste (Chen et al., 2014; Narukawa et al., 2010; Hayashi et al., 2010; Scharbert et al., 2004). The non-gallated catechins (C, EGC and EC) are less astringent and have a sweet aftertaste (Narukawa et al., 2010; Hara, 2001). Caffeine contributes greatly to the quality of tea, because of its sharp bitterness and briskness attributes (Obanda et al., 1997), while theanine has the sweet and umami taste (Golding et al., 2009; Kaneko et al., 2006).

In this study, there was a relationship between the major catechins (EGCG, ECG and EGC) and organoleptic results, indicating that they could be used to select high quality teas in a tea breeding programme. A strong positive correlation between EGCG and EGC, and taste implied that clones Clone55, Clone62, Clone71, TRFK Clone67, Clone48, Clone37 and Clone63 could be classified as high taste teas. Similarly, a strong positive correlation between mouthfeel and ECG, C, theogallin, GCG and GA implied that clones Clone97, Clone70, Clone53, Clone69, Clone56, Clone2 and CLONE101 could be categorized

as high mouthfeel teas. This was interesting considering Clone97, CLONE101 and Clone62 are commonly used by the tea industry (Kamunya et al., 2012) and at UTK (Gabriel Tuwei, Pers. com, June 2016, R&D UTK) as high quality standard checks. These findings agreed with Okinda Owuor et al. (2006), Owuor et al. (2007) and Obanda et al. (1997) who found that ECG, EGCG and EC correlated positively with organoleptic results from professional tea tasters in Kenyan teas. Similarly, in Central and Southern Africa (Wright et al., 2000a), and Sri Lanka , the quantities of EC and ECG in green leaves were shown to be reliable markers for identifying high-quality black tea. These findings were further confirmed by hierarchical cluster analysis. The compounds that were positively correlated with taste (EGCG, EGC and EC) were also clustered together. Similarly, the compounds that positively correlated with mouthfeel (ECG, C, theogallin, GA and GCG), also clustered together on the heatmap. According to their structural differences at the 5′ position in the B ring, tea catechins can be divided into dihydroxylated (ECG, EC and C) and trihydroxylated catechins (EGCG, EGC and GC) (Wei et al., 2018). EGC and EC are converted into esterified catechins (EGCG and ECG) via the sequential action of flavan-3-ol gallate synthase (FGS) (Liu et al., 2015). EGC and EGCG are formed on the same pathway. Similarly, ECG, EC and C are synthesized on the same pathway and this could explain their clustering on the heat map.

6.5. Conclusion

The levels of catechins, caffeine and theanine influence sensory tea quality, and can be exploited to select high quality varieties in a tea breeding programme. In this study, EGCG and EGC correlated with taste, and could be used to select tea varieties with high taste scores. Likewise, ECG correlated with mouthfeel and could therefore be used to select tea varieties with high mouthfeel. Results from this study are promising and suggest that it is possible to select high quality teas using only biochemical data. This will improve the accuracy, reproducibility, reduce the cost of analysis and reduce the subjectivity associated with organoleptic evaluation. Besides, analytical methods could be used early in the tea breeding programme, at the nursery stage to select desirable tea varieties, hence saving time of field planting and

miniature manufacture. However, only eight contrasting varieties were used for organoleptic evaluation and it will be important to validate these results using a large number of genetically diverse tea genotypes.

References

Bhattacharyya, N., Bandyopadhyay, R., Bhuyan, M., Tudu, B., Ghosh, D., Jana, A. (2008). Electronic nose for black tea classification and correlation of measurements with "Tea Taster" marks. *IEEE transactions on instrumentation and measurement, 57*(7), 1313-1321.

Chaturvedula, V. S. P., Prakash, I. (2011). The aroma, taste, color and bioactive constituents of tea. *Journal of Medicinal Plants Research, 5*(11), 2110-2124.

Chen, G. H., Yang, C. Y., Lee, S. J., Wu, C. C., Tzen, J. T. C. (2014). Catechin content and the degree of its galloylation in oolong tea are inversely correlated with cultivation altitude. *J Food Drug Anal, 22*(3), 303-309. doi:10.1016/j.jfda.2013.12.001

Chen, Q., Zhao, J., Liu, M., Cai, J., Liu, J. (2008). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *Journal of Pharmaceutical and Biomedical Analysis, 46*(3), 568-573.

Chen, Q., Zhao, J., Guo, Z., Wang, X. (2010). Determination of caffeine content and main catechins contents in green tea (Camellia sinensis L.) using taste sensor technique and multivariate calibration. *Journal of Food Composition and Analysis, 23*(4), 353-358. doi:10.1016/j.jfca.2009.12.010

Cheng, T. O. (2004). Will green tea be even better than black tea to increase coronary flow velocity reserve? *American Journal of Cardiology, 94*(9), 1223.

Corley, R. H. V., Tuwei, G. (2018). The Well-Bred Tea Bush. *Cambridge University Press.In: M. Carr (Author), Advances in Tea Agronomy*, 106-136.

El-Shahawi, M. S., Hamza, A., Bahaffi, S. O., Al-Sibaai, A. A., Abduljabbar, T. N. (2012). Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography. *Food Chem, 134*(4), 2268-2275. doi:10.1016/j.foodchem.2012.03.039

Fan, F.-Y., Shi, M., Nie, Y., Zhao, Y., Ye, J.-H., Liang, Y.-R. (2016). Differential behaviors of tea catechins under thermal processing: Formation of non-enzymatic oligomers. *Food Chemistry, 196*, 347-354.

Friedman, M., Kim, S. Y., Lee, S. J., Han, G. P., Han, J. S., Lee, K. R., Kozukue, N. (2005). Distribution of catechins, theaflavins, caffeine, and theobromine in 77 teas consumed in the United States. *Journal of food science, 70*(9), C550-C559.

Gallaher, R. N., Gallaher, K., Marshall, A. J., Marshall, A. C. (2006). Mineral analysis of ten types of commercially available tea. *Journal of Food Composition and Analysis, 19*, S53-S57.

Golding, J., Roach, P., Parks, S. (2009). *Production of high quality export green tea through integrated management*: RIRDC.

Hara, Y. (2001). *Green tea: health benefits and applications*: CRC press.

Harbowy, M. E., Balentine, D. A., Davies, A. P., Cai, Y. (1997). Tea chemistry. *Critical reviews in plant sciences, 16*(5), 415-480.

Hayashi, N., Chen, R., Hiraoka, M., Ujihara, T., Ikezaki, H. (2010). β-Cyclodextrin/surface plasmon resonance detection system for sensing bitter-astringent taste intensity of green tea catechins. *Journal of Agricultural and Food Chemistry, 58*(14), 8351-8356.

Ho, C.-T., Zheng, X., Li, S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9-27. doi:10.1016/j.fshw.2015.04.001

ITC. (2017). Annual Bulletin of Statistics. . from International Tea Committee.

Jain, S. M., Priyadarshan, P. M. (2009). *Breeding plantation tree crops: tropical species*: Springer.

Jeganathan, B., Kottawa-Arachchi, J. D., Ranatunga, M. A. B., Abeysinghe, I. S. B., Gunasekare, M. T. K., Bandara, B. M. R. (2017). Genotypic variation in biochemical compounds of the Sri Lankan tea

(Camellia sinensis L.) accessions and their relationships to quality and biotic stresses AU - Punyasiri, P. A. Nimal. *The Journal of Horticultural Science and Biotechnology, 92*(5), 502-512. doi:10.1080/14620316.2017.1289070

Kamunya, S. M., Wachira, F. N., Pathak, R. S., Muoki, R. C., Sharma, R. K. (2012). Tea improvement in Kenya. In *Global Tea Breeding* (pp. 177-226): Springer.

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

Kawakami, M. (1997). Comparison of extraction techniques for characterizing tea aroma and analysis of tea by GC-FTIR-MS. In *Plant Volatile Analysis* (pp. 211-229): Springer.

Koch, W., Kukula-Koch, W., Komsta, L., Marzec, Z., Szwerc, W., Glowniak, K. (2018). Green Tea Quality Evaluation Based on Its Catechins and Metals Composition in Combination with Chemometric Analysis. *Molecules, 23*(7). doi:10.3390/molecules23071689

Kumar, D., Gulati, A., Sharma, U. (2015). Determination of Theanine and Catechin in Camellia sinensis (Kangra Tea) Leaves by HPTLC and NMR Techniques. *Food Analytical Methods, 9*(6), 1666-1674. doi:10.1007/s12161-015-0343-z

Kumar, P. V. S., Basheer, S., Ravi, R., Thakur, M. S. (2011). Comparative assessment of tea quality by various analytical and sensory methods with emphasis on tea polyphenols. *Journal of food science and technology, 48*(4), 440-446.

Lee, J.E., Lee, B.-J., Chung, J.-O., Hwang, J.-A., Lee, S.-J., Lee, C.-H., Hong, Y.-S. (2010). Geographical and climatic dependencies of green tea (Camellia sinensis) metabolites: a 1H NMR-based metabolomics study. *Journal of Agricultural and Food Chemistry, 58*(19), 10582-10589.

Lee, J.E., Lee, B.-J., Chung, J.-O., Shin, H.-J., Lee, S.-J., Lee, C.-H., Hong, Y.-S. (2011). 1H NMR-based metabolomic characterization during green tea (Camellia sinensis) fermentation. *Food Research International, 44*(2), 597-604. doi:https://doi.org/10.1016/j.foodres.2010.12.004

Lee, J., Chambers, D. H. (2007). A lexicon for flavor descriptive analysis of green tea. *Journal of Sensory Studies, 22*(3), 256-272.

Lee, L. S., Kim, S. H., Kim, Y. B., Kim, Y. C. (2014). Quantitative analysis of major constituents in green tea with different plucking periods and their antioxidant activity. *Molecules, 19*(7), 9173-9186. doi:10.3390/molecules19079173

Liang, Y., Lu, J., Zhang, L., Wu, S., Wu, Y. (2003). Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions. *Food Chemistry, 80*(2), 283-290.

Magagna, F., Cordero, C., Cagliero, C., Liberto, E., Rubiolo, P., Sgorbini, B., Bicchi, C. (2017). Black tea volatiles fingerprinting by comprehensive two-dimensional gas chromatography - Mass spectrometry combined with high concentration capacity sample preparation techniques: Toward a fully automated sensomic assessment. *Food Chem, 225*, 276-287. doi:10.1016/j.foodchem.2017.01.003

Mahindroo, N. (2000). Tea-Camellia sinensis-report. *ICS-UNIDO, Trieste, Italy*.

Mondal, T. K. (2014). Genetics and Breeding. In *Breeding and Biotechnology of Tea and its Wild Species* (pp. 9-34).

Mukhtar, H., Ahmad, N. (2000). Tea polyphenols: prevention of cancer and optimizing health–. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.

Nakabayashi, T. (1994). Chemical components in tea leaves. *Chemistry and functions of green, black and Oolong teas*, 20-42.

Narukawa, M., Kimata, H., Noga, C., Watanabe, T. (2010). Taste characterisation of green tea catechins. *International Journal of Food Science & Technology, 45*(8), 1579-1585. doi:10.1111/j.1365-2621.2010.02304.x

Narukawa, M., Noga, C., Ueno, Y., Sato, T., Misaka, T., Watanabe, T. (2011). Evaluation of the bitterness of green tea catechins by a cell-based assay with the human bitter taste receptor hTAS2R39. *Biochemical and biophysical research communications, 405*(4), 620-625.

Obanda, M., Owuor, P. O., Taylor, S. J. (1997). Flavanol composition and caffeine content of green leaf as quality potential indicators of Kenyan black teas. *Journal of the Science of Food and Agriculture, 74*(2), 209-215.

Okinda Owuor, P., Obanda, M., Nyirenda, H. E., Mphangwe, N. I. K., Wright, L. P., Apostolides, Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (Camellia sinensis) produced in Kenya and comparison with similar teas from Malawi and South Africa. *Food Chemistry, 97*(4), 644-653. doi:10.1016/j.foodchem.2005.04.027

Owuor, P. O., Obanda, M. (2007). The use of green tea (Camellia sinensis) leaf flavan-3-ol composition in predicting plain black tea quality potential. *Food Chemistry, 100*(3), 873-884. doi:10.1016/j.foodchem.2005.10.030

Payne, R. W. (2015). *Introduction to GenStat® for windows™*. Retrieved from Hemel Hempstead:

Samynathan, R., Perisamy, S. k., Gandhi, S., Anitha, J., Sanmugam, G., Padmanabhan, M., Velliyur Kanniappan, G. (2016). Biochemical and molecular analysis of Camellia sinensis (L.) O. Kuntze tea from the selected P/11/15 clone. *Journal of Genetic Engineering and Biotechnology, 14*(1), 69-75. doi:https://doi.org/10.1016/j.jgeb.2015.12.004

Scharbert, S., Hofmann, T. (2005). Molecular definition of black tea taste by means of quantitative studies, taste reconstitution, and omission experiments. *Journal of Agricultural and Food Chemistry, 53*(13), 5377-5384.

Scharbert, S., Holzmann, N., Hofmann, T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *Journal of Agricultural and Food Chemistry, 52*(11), 3498-3508.

TBK. (2017). Tea industry performance in 2017. from The Tea Board of Kenya.

Tu, Y., Bian, M., Wan, Y., Fei, T. (2018). Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. *PeerJ, 6*, e4858.

Wang, K., Liu, F., Liu, Z., Huang, J., Xu, Z., Li, Y., Chen, J., Gong, Y., Yang, X. (2010). Analysis of chemical components in oolong tea in relation to perceived quality. *International Journal of Food Science & Technology, 45*(5), 913-920. doi:10.1111/j.1365-2621.2010.02224.x

Wright, L. P., Mphangwe, N. I. K., Nyirenda, H. E., Apostolides, Z. (2000). Analysis of caffeine and flavan-3-ol composition in the fresh leaf of Camellia sinesis for predicting the quality of the black tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 80*(13), 1823-1830. doi:doi:10.1002/1097-0010(200010)80:13<1823::AID-JSFA702>3.0.CO;2-E

Yamamoto, T., Juneja, L. R., Chu, D. C., Kim, M. (1997). Chemistry and Applications of Green Tea CRC press. *Florida, USA*.

Yamanishi, T. (1995). Special issue on tea: Flavor of tea. *Food Review Int., 11*, 477-525.

Yang, Z., Baldermann, S., Watanabe, N. (2013). Recent studies of the volatile compounds in tea. *Food Research International, 53*(2), 585-599.

Yu, P., Yeo, A. S., Low, M. Y., Zhou, W. (2014). Identifying key non-volatile compounds in ready-to-drink green tea and their impact on taste profile. *Food Chem, 155*, 9-16. doi:10.1016/j.foodchem.2014.01.046

Yuan, Y., Song, Y., Jing, W., Wang, Y., Yang, X., Liu, D. (2014). Simultaneous determination of caffeine, gallic acid, theanine,(−)-epigallocatechin and (−)-epigallocatechin-3-gallate in green tea using quantitative 1 H-NMR spectroscopy. *Analytical Methods, 6*(3), 907-914.

Zheng, X.-Q., Li, Q.-S., Xiang, L.-P., Liang, Y.-R. (2016). Recent advances in volatiles of teas. *Molecules, 21*(3), 338.

Table S6. 1: List of genotypes and their crosses analyzed used using NMR spectroscopy.

	Clone	Female	Male	Species
1	Clone1	502/7	CLONE98	*C. sinensis*
2	Clone2	CLONE98	15/10	*C. sinensis*
3	Clone3	SFS150	CLONE100	*C. sinensis*
4	Clone4	CLONE98	15/10	*C. sinensis*
5	Clone5	CLONE98	CLONE100	*C. sinensis*
6	Clone6	CLONE98	CLONE100	*C. sinensis*
7	Clone7	CLONE98	CLONE100	*C. sinensis*
8	Clone8	CLONE98	CLONE100	*C. sinensis*
9	Clone9	CLONE101	CLONE98	*C. sinensis*
10	Clone10	PC80	CLONE100	*C. sinensis*
11	Clone11	SFS150	CLONE100	*C. sinensis*
12	Clone12	CLONE100	PC80	*C. sinensis*
13	Clone13	CLONE104	CLONE100	*C. sinensis*
14	Clone14	SFS150	CLONE100	*C. sinensis*
15	Clone15	CLONE104	CLONE98	*C. sinensis*
16	Clone16	SFS150	CLONE98	*C. sinensis*
17	Clone17	PC110	CLONE98	*C. sinensis*
18	Clone18	SFS150	CLONE98	*C. sinensis*
19	Clone19	SFS150	CLONE98	*C. sinensis*
20	Clone20	PC110	CLONE98	*C. sinensis*
21	Clone21	PC110	CLONE98	*C. sinensis*
22	Clone22	CLONE104	CLONE98	*C. sinensis*
23	Clone23	PC110	CLONE98	*C. sinensis*
24	Clone24	CLONE98	CLONE97	*C. sinensis*
25	Clone25	PC87	CLONE98	*C. sinensis*
26	Clone26	CLONE104	CLONE98	*C. sinensis*
27	Clone27	SFS150	CLONE98	*C. sinensis*
28	Clone28	SFS204	CLONE98	*C. sinensis*
29	Clone29	SFS150	15/10	*C. sinensis*
30	Clone30	CLONE104	15/10	*C. sinensis*
31	Clone31	CLONE101	CLONE98	*C. sinensis*
32	Clone32	BBT207	15/10	*C. sinensis*
33	Clone33	BBT207	CLONE98	*C. sinensis*
34	Clone34	BBT207	CLONE97	*C. sinensis*
35	Clone35	864	CLONE98	*C. sinensis*
36	Clone36	CLONE98	CLONE101	*C. sinensis*
37	Clone37	Clone65	CLONE100	*C. sinensis*
38	Clone38	18233	CLONE100	*C. sinensis*
39	Clone39	864	CLONE98	*C. sinensis*
40	Clone40	12/108	CLONE98	*C. sinensis*

41	Clone41	CLONE98	15/10	*C. sinensis*
42	Clone42	14/6	PC110	*C. sinensis*
43	Clone43	12/108	15/10	*C. sinensis*
44	Clone44	14/6	CLONE98	*C. sinensis*
45	Clone45	CLONE98	PC110	*C. sinensis*
46	Clone46	CLONE98	PC110	*C. sinensis*
47	Clone47	CLONE95	CLONE98	*C. sinensis*
48	Clone48	14/3	42104	*C. sinensis*
49	Clone49	42104	CLONE98	*C. sinensis*
50	‘Clone50	502/7	864	*C. sinensis*
51	Clone51	SFS150	42163	*C. sinensis*
52	Clone52	SFS150	68	*C. sinensis*
53	Clone53	SFS150	42280	*C. sinensis*
54	Clone54	1607/2	CLONE98	*C. sinensis*
55	Clone55	1602/7	15/10	*C. sinensis*
56	Clone56	Clone68	CLONE98	*C. sinensis*
57	Clone57	803/17	PC110	*C. sinensis*
58	Clone58	CLONE101	PC110	*C. sinensis*
59	Clone59	864	12/108	*C. sinensis*
60	Clone60			
61	Clone61	NA	NA	*C. sinensis*
62	Clone62			
63	Clone63	NA	NA	*C. sinensis*
64	Clone64	CLONE95	BB152	*C. sinensis*
65	Clone65	NA	NA	*C. sinensis*
66	Clone66	NA	NA	*C. sinensis*
67	Clone67	NA	NA	*C. sinensis*
68	Clone68	CLONE95	BB5	*C. sinensis*
69	Clone69	CLONE101	CLONE97	*C. sinensis*
70	Clone70	PC110	CLONE97	*C. sinensis*
71	Clone71	SFS150	12/108	*C. sinensis*
72	Clone72	PC87	15/10	*C. sinensis*
73	Clone73	BBT207	CLONE100	*C. sinensis*
74	Clone74	PC87	15/10	*C. sinensis*
75	Clone75	1/157	CLONE98	*C. sinensis*
76	Clone76	42163	CLONE97	*C. sinensis*
77	Clone77	CLONE104	29221	*C. sinensis*
78	Clone78	C12	Clone67	*C. sinensis*
79	Clone79	CLONE99	CLONE100	*C. sinensis*
80	Clone80	864	CLONE98	*C. sinensis*
81	Clone81	C12	CLONE98	*C. sinensis*
82	Clone82	C12	CLONE98	*C. sinensis*
83	Clone83	C12	CLONE101	*C. sinensis*

84	Clone84	C12	CLONE99	*C. sinensis*
85	Clone85	C12	42163	*C. sinensis*
86	Clone86	C12	CLONE101	*C. sinensis*
87	Clone87	C12	CLONE101	*C. sinensis*
88	Clone88	C12	CLONE99	*C. sinensis*
89	Clone89	C12	CLONE99	*C. sinensis*
90	Clone90	CLONE99	42163	*C. sinensis*
91	Clone91	CLONE98	15/10	*C. sinensis*
92	Clone92	PC110	CLONE101	*C. sinensis*
93	Clone93	14/3	42104	*C. sinensis*
94	Clone94	CLONE98	PC110	*C. sinensis*
95	CLONE95			
96	CLONE96	NA	NA	*C. sinensis*
97	CLONE97	NA	NA	*C. sinensis*
98	CLONE98	NA	NA	*C. sinensis*
99	CLONE99	NA	NA	*C. sinensis*
100	CLONE100	NA	NA	*C. sinensis*
101	CLONE101	NA	NA	*C. sinensis*
102	CLONE102	NA	NA	*C. sinensis*
103	CLONE103	NA	NA	*C. sinensis*
104	CLONE104	NA	NA	*C. sinensis*
105	CLONE105	NA	NA	*C. sinensis*
106	CLONE106	NA	NA	*C. sinensis*

Table S6. 2: Mean biochemical traits across Jamji and Kericho. The values colored green indicates not significantly ($p<0.05$) different from each other for each trait.

	clone	TB	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
1	CLONE97	3.52	37.49	24.39	4.85	25.00	24.71	112.61	20.67	72.75	11.76	57.40	21.76
2	CLONE101	4.98	36.06	20.71	5.29	30.82	25.54	81.31	46.86	91.77	8.57	39.90	13.86
3	Clone69	3.98	39.88	22.38	5.73	31.57	28.11	89.62	40.77	90.22	5.84	34.59	22.11
4	Clone70	2.57	35.48	21.73	5.82	28.41	33.01	97.77	30.59	68.34	8.76	50.85	14.18
5	Clone37	3.55	39.17	18.96	4.85	34.67	22.20	65.05	51.35	106.51	5.46	31.81	13.29
6	Clone89	4.69	38.40	15.38	4.13	32.53	20.91	59.85	53.05	101.46	7.52	39.38	20.44
7	Clone79	4.51	36.72	16.55	4.58	31.11	21.74	62.95	53.73	97.87	4.47	40.89	21.25
8	Clone61	6.84	41.59	17.96	3.54	29.17	16.15	71.80	25.78	93.42	12.78	63.28	14.41
9	Clone25	2.10	29.67	14.68	5.06	28.76	27.02	64.14	69.30	77.97	5.00	39.44	15.34
10	Clone55	5.78	44.18	16.39	4.21	34.95	17.99	60.22	40.04	120.82	4.25	33.21	15.07
11	Clone57	5.82	41.38	14.43	4.66	30.38	23.92	64.42	56.30	92.37	4.49	38.60	19.66
12	Clone68	6.24	40.35	18.25	4.07	34.47	18.59	71.77	36.16	102.91	4.37	41.47	21.34
13	Clone87	4.81	42.47	15.69	4.47	29.87	23.41	70.88	50.06	85.22	3.75	44.39	20.38
14	Clone81	5.52	40.40	15.32	3.36	24.99	18.29	72.88	39.73	75.46	12.82	63.10	29.82
15	Clone63	6.17	36.94	14.79	4.51	32.59	19.73	57.12	50.69	106.27	4.89	35.86	18.46
16	Clone53	3.02	33.57	21.40	4.98	28.38	24.27	93.17	26.02	74.71	8.54	48.94	21.18

17	CLONE95	6.70	41.15	18.11	4.52	31.65	20.80	61.03	46.69	101.11	6.27	35.82	21.54
18	Clone62	4.51	42.73	15.43	4.55	33.35	22.13	54.82	52.07	108.96	3.47	28.01	17.18
19	CLONE102	6.49	42.48	15.24	4.76	29.91	22.13	58.27	49.12	103.06	4.53	35.67	15.60
20	Clone56	5.85	37.17	18.91	3.66	28.67	16.74	85.83	17.13	91.86	10.89	51.40	19.81
21	Clone71	3.96	43.40	15.37	4.32	31.14	20.51	55.24	51.08	107.86	3.31	33.10	16.70
22	Clone60	5.40	37.17	15.78	4.30	30.81	20.08	55.63	49.92	102.68	5.75	35.95	18.66
23	Clone46	3.93	42.06	14.60	4.27	29.76	22.20	62.02	51.28	93.39	5.72	35.68	18.87
24	Clone9	4.21	37.69	16.80	3.83	25.67	20.21	75.69	37.84	77.76	9.67	53.11	20.46
25	Clone3	3.47	35.98	15.26	3.72	32.43	18.26	58.25	50.80	103.52	5.55	30.50	14.31
26	Clone67	4.69	42.11	13.75	3.98	29.32	17.18	50.26	49.28	106.93	7.19	38.57	24.64
27	Clone18	3.72	38.69	16.18	4.39	31.37	23.06	57.63	58.73	93.61	5.93	28.01	20.76
28	Clone21	3.16	36.87	15.66	3.58	25.13	19.20	73.91	35.65	77.41	11.04	55.87	19.73
29	Clone29	3.29	35.90	15.09	4.91	32.52	22.36	50.95	62.96	99.84	2.93	26.27	14.97
30	Clone91	3.20	34.28	18.59	4.30	31.24	21.92	69.85	39.70	93.64	4.68	35.47	22.72
31	Clone31	5.75	33.49	18.41	3.47	26.62	17.58	75.49	33.20	74.42	11.31	56.46	20.45
32	Clone49	2.72	44.02	14.85	4.30	31.62	21.59	60.95	47.57	99.29	7.26	26.74	18.57
33	Clone92	6.21	43.00	18.99	5.33	32.97	26.90	65.73	47.57	91.24	4.62	25.49	13.89
34	Clone64	5.58	42.79	15.03	3.99	30.87	18.16	59.64	40.53	96.60	6.45	41.52	14.03
35	CLONE100	4.13	37.52	15.17	3.89	30.61	17.42	53.83	47.72	101.45	3.41	39.23	20.40
36	Clone66	3.77	36.97	14.40	4.65	32.72	21.04	48.59	57.72	99.30	4.68	28.41	10.90
37	Clone51	4.38	40.08	17.30	4.51	32.37	21.69	61.43	41.18	91.68	5.27	37.01	18.89
38	Clone17	4.05	37.46	16.22	3.47	26.12	18.13	68.94	35.53	76.51	9.22	56.09	22.37
39	Clone65	2.87	33.69	13.93	3.69	32.10	19.46	46.39	58.87	100.58	4.89	26.94	10.70
40	Clone48	5.70	46.66	17.55	4.03	33.47	19.62	53.86	42.32	106.72	3.79	29.34	14.83
41	Clone82	3.42	36.08	17.52	4.48	31.29	22.59	63.96	48.06	89.71	4.47	28.71	21.00
42	Clone72	4.12	28.36	16.50	5.14	31.02	26.16	55.79	64.42	82.80	4.03	23.97	12.09
43	Clone16	3.15	43.07	16.70	3.96	28.05	19.25	73.60	29.22	84.27	7.39	46.35	21.49
44	Clone94	2.88	36.70	15.52	4.52	26.57	23.35	75.72	32.71	69.43	8.64	51.45	13.94
45	Clone30	4.53	38.41	13.21	3.45	28.96	16.44	49.71	51.18	94.93	5.43	39.77	16.17
46	Clone33	4.11	38.00	13.15	4.23	29.40	21.90	54.58	50.61	90.09	4.65	34.47	16.56
47	CLONE103	3.14	33.63	13.33	4.18	30.92	20.20	47.10	58.94	89.81	4.34	33.67	17.47
48	Clone5	3.69	38.08	16.85	3.50	24.33	16.69	77.42	25.16	73.80	7.42	59.73	17.01
49	Clone86	6.26	42.63	17.16	4.98	28.95	25.23	64.71	45.55	81.23	6.56	31.62	26.21
50	Clone13	3.94	37.20	12.90	4.03	30.17	16.79	44.71	59.17	92.73	5.30	34.57	17.31
51	Clone34	3.66	37.99	18.14	4.39	25.09	19.97	68.56	35.12	73.25	7.80	53.59	20.43
52	Clone2	3.47	34.75	18.83	3.76	23.73	17.78	84.34	22.72	66.08	10.45	58.17	14.46
53	Clone90	4.12	41.17	14.16	3.91	31.32	18.96	49.51	49.52	97.52	5.46	30.83	16.56
54	Clone85	4.32	33.89	14.12	3.14	31.30	16.05	40.69	60.15	101.46	3.71	29.61	19.18
55	CLONE98	4.79	39.95	16.59	3.37	25.62	17.44	74.42	26.95	79.13	7.76	51.28	19.13
56	Clone14	2.88	36.22	13.03	3.82	28.77	18.89	49.54	57.70	92.98	3.91	30.05	16.95
57	Clone44	4.10	35.74	15.98	3.36	27.76	19.29	63.87	38.74	82.88	6.34	41.28	25.48
58	Clone8	4.77	38.99	17.07	3.34	28.53	15.86	64.85	26.79	91.82	7.64	43.94	18.78
59	Clone54	6.36	39.41	16.23	3.79	29.63	18.68	60.51	36.67	93.61	6.36	33.63	18.53
60	Clone36	3.06	34.39	16.05	4.03	26.01	19.83	70.75	34.00	76.65	7.20	44.22	18.97
61	Clone19	3.17	38.32	13.57	3.89	28.47	18.58	54.81	48.60	86.22	5.88	36.07	18.42
62	Clone40	4.66	39.64	19.03	3.61	28.54	18.21	71.95	27.30	83.04	7.29	41.33	17.92
63	Clone38	4.40	37.18	14.06	3.58	29.12	16.69	53.26	43.13	94.60	6.10	34.77	21.42
64	Clone7	4.22	36.08	17.07	2.99	26.03	14.76	67.07	26.58	84.42	9.19	49.55	19.27
65	Clone83	4.86	33.84	14.15	3.59	27.85	18.89	54.69	50.16	86.10	6.15	33.35	23.57
66	Clone32	3.62	42.31	15.99	4.08	29.92	18.43	49.31	49.21	94.21	5.45	30.04	15.56

67	Clone42	6.33	40.71	12.85	3.88	27.18	19.17	46.59	55.58	89.45	6.18	32.41	20.11
68	Clone12	3.88	37.88	17.81	4.55	30.47	21.54	57.59	42.67	94.86	3.83	25.46	20.36
69	Clone50	6.10	37.46	17.10	4.36	27.93	21.55	57.68	47.18	87.40	4.00	30.43	16.82
70	Clone78	4.94	37.06	14.21	3.69	26.73	20.59	54.83	50.90	81.68	6.71	34.64	25.80
71	Clone22	4.46	41.78	15.96	2.94	27.46	13.93	59.23	27.37	91.94	8.37	47.36	24.72
72	Clone75	4.25	41.46	12.48	3.45	27.14	17.21	50.90	48.38	79.68	7.31	44.99	26.71
73	Clone93	4.50	39.67	15.45	3.92	27.46	20.75	62.03	46.49	72.85	7.01	38.27	18.43
74	Clone4	5.60	42.46	16.83	3.67	27.18	16.72	62.90	28.64	82.14	9.20	47.47	17.57
75	Clone28	5.04	43.06	16.80	3.80	26.61	19.16	75.24	23.39	73.45	7.95	48.09	23.33
76	Clone1	5.67	38.04	16.77	4.02	29.95	17.71	53.38	43.84	91.29	4.77	32.70	18.64
77	Clone76	3.82	33.75	16.59	4.64	30.98	24.37	57.48	51.11	74.90	3.96	30.24	15.65
78	Clone6	3.08	34.49	13.24	3.90	26.88	19.64	52.05	51.36	84.08	5.51	32.43	16.02
79	Clone35	4.84	40.24	15.45	3.67	28.57	17.33	62.86	33.91	79.16	6.16	43.61	18.02
80	Clone58	4.52	33.22	16.27	4.59	27.59	23.32	61.65	41.78	77.49	6.31	31.98	20.00
81	Clone27	3.44	39.43	12.63	3.88	29.74	18.32	43.98	54.63	97.16	3.89	22.36	19.39
82	CLONE99	3.63	34.34	13.24	4.24	27.68	20.78	48.21	52.26	86.77	4.35	29.92	17.13
83	Clone73	4.40	37.00	14.69	4.46	28.60	20.68	48.57	51.51	89.03	3.93	27.51	21.72
84	Clone45	2.96	33.35	11.64	3.84	22.87	19.46	59.12	39.03	59.89	9.11	60.17	16.96
85	Clone23	2.20	34.67	15.48	3.69	22.83	18.85	72.18	24.92	67.55	9.64	51.13	20.81
86	Clone20	2.60	32.88	12.99	5.66	28.14	27.74	54.13	52.13	67.11	5.06	31.06	19.95
87	Clone52	3.71	41.51	13.22	4.05	28.42	18.11	46.66	51.11	84.13	4.78	31.71	16.77
88	Clone84	5.31	35.40	14.22	4.06	31.70	18.26	45.02	48.00	94.25	3.47	24.02	29.10
89	Clone88	4.45	37.98	12.43	4.32	28.29	22.58	44.38	65.02	70.83	4.03	28.93	17.89
90	Clone39	2.53	35.93	17.56	3.12	25.24	14.37	68.35	18.53	83.96	6.84	46.58	12.23
91	Clone10	2.68	27.92	14.87	3.79	28.13	18.32	47.74	48.00	90.35	3.24	25.91	16.77
92	Clone26	2.44	30.84	15.47	3.80	27.44	17.91	47.37	47.56	86.40	4.16	30.20	15.05
93	Clone43	3.58	36.58	16.83	3.50	25.61	16.27	60.75	29.58	74.48	8.34	45.06	22.49
94	Clone59	6.13	39.54	17.99	3.63	27.92	17.99	59.50	30.46	87.39	4.57	31.45	19.25
95	Clone77	3.86	35.24	14.02	3.29	27.30	15.54	47.83	42.63	86.61	5.91	31.65	17.95
96	Clone24	3.48	35.61	11.47	4.72	25.01	24.12	48.65	57.54	68.92	4.83	26.99	17.30
97	Clone47	5.62	40.19	15.46	3.78	28.07	18.47	49.03	40.65	87.92	3.37	27.80	22.32
98	Clone41	3.97	39.36	18.03	3.38	27.20	15.64	63.69	23.76	77.06	6.15	40.76	25.69
99	Clone80	6.51	40.77	15.79	3.55	28.66	17.02	45.40	38.31	94.13	4.29	22.70	27.11
100	Clone74	3.83	34.15	15.03	4.35	26.75	20.95	45.47	51.77	72.73	4.12	26.51	26.23
101	Clone11	2.41	35.11	12.23	3.34	25.28	16.08	40.51	49.83	80.27	4.10	27.12	28.85
102	Clone15	3.44	37.61	12.88	3.04	22.51	13.65	53.91	26.95	66.99	7.10	44.47	21.71
103	CLONE96	2.85	30.37	11.67	4.27	24.21	22.04	46.74	41.59	66.51	4.88	26.85	13.70

Table S6. 3: Mean biochemical traits for Jamji. The values colored green indicates not significantly ($p<0.05$) different from each other for each trait.

	clone	Theobromine	Caffeine	Theogallin	GA	Epicatechin	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
1	Clone79	6.53	43.17	19.96	5.37	36.15	25.08	77.59	58.12	114.70	4.69	50.75	24.52
2	Clone56	7.85	42.44	22.36	4.59	34.49	19.48	98.24	19.98	107.54	13.84	57.29	19.19
3	Clone81	7.06	42.61	17.33	3.79	26.85	19.56	80.16	38.65	84.72	15.53	65.29	30.36
4	CLONE102	6.87	44.83	15.38	5.41	32.96	23.62	64.46	53.30	110.03	5.64	39.73	12.91
5	Clone70	1.93	35.74	19.78	3.91	25.90	21.78	101.88	22.02	73.79	12.30	69.01	13.61

6	Clone25	2.68	28.58	15.07	5.21	31.01	27.63	66.14	73.97	82.53	5.52	39.46	16.80
7	CLONE97	3.45	37.04	24.70	4.71	25.57	23.74	112.16	21.52	73.38	12.89	55.63	22.43
8	Clone69	4.86	39.78	24.66	6.11	33.48	28.77	92.50	39.32	91.44	5.01	30.02	24.07
9	Clone89	5.05	38.36	15.39	4.25	32.17	21.35	62.69	52.95	101.32	7.67	41.86	20.21
10	Clone71	4.44	46.97	16.20	4.88	34.86	21.30	59.60	51.50	114.16	3.31	32.63	19.38
11	Clone92	7.26	46.40	21.88	5.99	37.27	28.52	72.78	48.11	100.82	4.89	24.34	11.54
12	Clone37	4.41	41.50	20.68	5.24	35.41	22.29	68.63	46.04	105.51	5.55	31.94	18.03
13	CLONE101	5.63	37.70	21.45	5.67	31.73	25.07	79.89	43.09	91.95	6.42	36.17	15.76
14	Clone53	3.19	37.06	20.48	5.26	29.75	24.24	86.97	30.43	84.31	8.42	46.27	22.65
15	Clone57	6.29	42.18	14.39	4.38	32.16	22.00	65.98	52.62	94.22	3.47	38.64	18.27
16	Clone87	5.96	41.69	16.35	4.77	29.68	24.53	74.59	48.63	83.31	3.68	44.22	17.92
17	Clone21	2.93	35.91	16.53	3.55	26.82	19.02	78.20	35.22	83.19	11.23	54.23	18.25
18	Clone66	3.62	37.22	15.40	5.25	33.53	21.44	48.49	59.73	111.66	2.96	29.94	9.59
19	Clone68	6.97	41.61	18.48	4.67	35.29	19.06	68.04	37.24	106.53	3.72	37.28	25.09
20	TRFK Clone67	5.01	44.71	14.28	4.39	33.42	17.73	48.12	52.17	115.85	7.79	29.82	27.52
21	Clone55	5.68	44.28	15.85	4.16	32.91	16.88	61.84	41.16	113.05	4.36	34.41	15.12
22	Clone46	3.57	41.98	15.31	4.46	30.01	22.75	65.38	48.07	96.50	6.20	35.05	14.74
23	Clone90	4.45	42.20	14.71	4.29	32.22	19.69	53.17	52.17	106.65	5.52	33.12	14.96
24	Clone6	4.54	37.47	15.93	4.73	32.02	22.29	60.52	53.14	96.15	7.11	31.20	17.64
25	Clone3	4.02	35.25	16.78	4.21	32.96	19.61	63.26	47.50	103.59	4.82	29.30	14.55
26	Clone49	2.64	44.42	14.56	4.40	30.38	21.16	62.29	50.06	97.74	7.77	30.64	18.20
27	Clone63	6.09	38.60	14.50	4.67	32.33	18.65	52.49	50.82	107.31	5.08	33.03	24.66
28	Clone31	6.66	34.67	19.92	3.92	27.64	18.78	76.44	33.97	76.05	14.01	52.42	24.78
29	Clone61	7.51	41.68	18.60	4.04	30.20	16.35	65.68	24.65	91.34	13.92	56.43	17.19
30	Clone18	4.51	38.70	16.90	4.69	30.81	24.04	56.78	57.51	95.22	6.57	27.37	26.75
31	CLONE100	5.16	38.03	16.84	4.49	33.58	18.73	54.71	49.06	103.10	3.09	35.94	25.43
32	Clone17	5.02	37.07	18.59	3.71	28.69	18.16	72.40	32.07	85.79	4.06	56.23	22.00
33	CLONE95	7.89	42.48	18.49	4.60	30.33	20.25	59.60	45.36	99.87	7.21	34.71	23.66
34	CLONE98	5.42	41.47	18.04	3.47	28.47	17.76	77.53	28.22	86.31	8.56	49.95	20.75
35	Clone30	5.16	39.99	14.13	3.71	30.70	17.69	50.85	52.39	101.50	5.19	38.31	21.45
36	CLONE105	5.29	32.02	14.83	4.05	31.86	19.33	48.60	62.63	97.84	5.28	30.99	19.60
37	Clone38	4.65	38.41	16.12	3.97	31.10	17.26	58.83	42.17	101.88	7.74	37.24	17.43
38	Clone48	6.63	47.06	19.53	4.53	34.26	20.96	57.45	42.81	103.80	4.71	30.88	14.19
39	Clone54	7.69	41.56	17.89	4.16	32.11	19.51	66.43	36.99	98.71	6.43	34.56	18.00
40	Clone91	3.42	34.96	18.55	4.33	29.96	20.62	69.60	37.94	94.88	3.73	37.95	24.05
41	Clone2	3.57	36.34	19.10	4.17	26.62	18.37	86.24	23.14	69.93	10.13	60.05	16.65
42	Clone9	4.08	40.11	16.35	3.90	27.10	19.68	72.66	38.57	80.70	8.23	47.28	22.54
43	Clone60	5.78	39.57	15.48	3.90	32.62	17.25	51.31	51.38	102.52	5.66	33.16	22.84
44	Clone13	4.21	37.28	13.59	4.28	31.93	17.01	46.12	62.03	100.59	3.27	32.92	19.39
45	Clone62	5.39	46.10	16.07	4.77	30.92	22.24	53.86	46.54	107.61	3.59	28.93	17.46
46	Clone94	3.14	38.70	15.84	4.66	26.54	23.41	77.72	32.38	68.69	11.62	52.47	14.96
47	Clone86	7.12	42.05	18.50	5.38	29.34	26.42	68.11	44.33	85.46	6.26	31.66	22.91
48	CLONE103	3.87	33.91	15.09	4.61	31.92	21.53	49.46	57.44	92.34	4.30	32.19	14.39
49	Clone14	3.56	35.57	14.24	4.12	30.37	19.67	52.03	55.29	98.14	3.88	28.90	18.43
50	Clone27	4.17	39.55	13.93	4.26	32.15	19.56	47.62	56.33	102.82	5.44	24.14	20.75
51	Clone40	5.43	40.73	20.26	3.66	30.00	17.28	75.72	26.03	88.04	6.71	44.26	18.34
52	Clone64	5.62	42.56	15.71	4.43	30.96	17.76	56.14	38.75	104.15	4.20	36.07	13.78
53	Clone16	3.36	42.14	17.28	4.09	28.39	19.19	74.01	29.58	83.84	7.82	44.57	22.03
54	Clone50	7.49	38.93	17.41	4.72	27.62	22.24	59.66	48.64	89.83	4.55	34.11	16.69
55	Clone34	4.36	39.71	18.82	4.78	27.79	21.11	70.58	36.78	70.55	6.52	53.03	24.67

56	Clone29	3.03	35.91	16.06	5.12	31.53	22.26	49.50	62.10	98.16	2.83	19.96	14.78
57	Clone5	4.61	38.15	17.32	3.75	25.94	17.04	77.50	24.57	77.02	6.22	57.29	19.29
58	Clone78	5.50	38.41	15.39	3.80	27.06	21.06	59.80	44.94	85.82	7.54	39.34	24.17
59	Clone82	3.91	34.05	18.93	4.54	29.61	21.72	69.11	42.84	85.03	5.29	31.16	20.80
60	Clone73	5.70	37.73	17.24	5.04	30.24	22.59	53.91	48.70	95.94	4.25	28.71	18.88
61	Clone12	3.64	36.64	17.34	4.16	31.63	19.74	59.65	42.49	101.07	4.41	25.36	20.18
62	Clone75	5.38	46.64	13.81	4.01	27.94	19.03	55.36	47.91	77.73	6.66	49.01	31.02
63	Clone65	3.50	36.13	14.65	3.76	32.47	18.77	44.90	54.06	103.74	4.51	24.28	12.99
64	Clone8	5.77	41.37	17.76	3.71	28.45	16.37	65.82	25.55	93.45	9.34	41.90	22.23
65	Clone85	5.21	32.80	14.53	3.13	30.48	15.98	41.41	59.50	100.62	3.05	29.27	17.32
66	Clone39	2.50	38.79	18.95	3.40	26.33	16.05	75.44	19.14	84.23	5.45	53.46	12.66
67	Clone72	5.09	27.76	17.21	5.17	31.20	25.71	55.72	63.64	75.88	3.39	23.45	15.19
68	Clone22	4.67	41.92	16.75	3.14	28.01	14.41	60.93	27.62	92.70	7.91	47.31	29.60
69	Clone42	7.04	41.80	14.10	3.95	27.78	18.99	49.98	51.95	90.86	5.74	33.61	19.12
70	Clone76	4.80	33.98	18.25	5.02	32.41	25.62	56.39	55.07	76.19	4.28	28.69	16.27
71	Clone41	5.24	41.51	20.04	3.97	29.24	16.60	70.55	25.12	83.85	7.90	44.90	29.86
72	Clone83	4.36	33.23	13.57	3.31	26.53	18.14	55.33	48.89	84.92	8.13	35.27	24.94
73	Clone88	5.10	39.46	12.87	4.49	28.92	23.46	46.48	68.95	74.09	3.87	30.47	18.02
74	Clone36	2.95	33.97	16.90	4.03	25.30	19.11	69.25	34.18	79.25	6.15	42.08	21.44
75	Clone19	4.42	40.15	14.65	4.25	29.21	19.30	53.33	45.12	90.73	5.74	31.52	24.87
76	Clone51	4.16	40.24	16.89	4.47	30.01	19.81	59.10	36.71	87.08	3.28	38.39	19.89
77	Clone32	4.41	44.02	17.37	4.36	29.81	19.03	48.18	47.30	98.29	4.80	26.06	16.74
78	Clone52	4.37	41.55	14.23	4.23	29.19	18.79	51.22	46.20	87.29	4.38	35.83	15.31
79	Clone43	3.34	40.45	19.09	4.11	25.83	18.33	64.99	31.57	73.21	8.99	49.95	18.55
80	Clone4	6.58	42.32	17.65	4.08	29.26	17.62	63.75	28.03	85.58	5.79	42.61	16.76
81	Clone28	5.10	43.73	17.21	3.88	26.34	18.68	75.65	21.63	74.37	8.22	47.21	25.46
82	Clone84	5.91	34.48	15.55	4.57	31.75	18.87	46.22	48.62	99.35	3.39	22.56	30.23
83	Clone1	6.52	38.02	17.68	4.06	29.32	18.31	54.76	40.23	91.03	4.30	32.67	16.04
84	Clone10	2.57	28.83	16.05	4.18	29.78	19.01	49.45	48.89	95.59	3.18	23.86	20.26
85	Clone20	1.95	31.82	12.69	5.49	28.41	27.15	54.21	52.97	71.48	5.29	28.72	21.83
86	Clone44	3.77	37.59	15.18	3.24	26.56	18.31	64.53	30.49	77.16	5.46	45.68	24.28
87	Clone59	6.68	39.24	19.16	3.72	29.29	18.05	63.69	28.40	91.26	4.46	32.74	16.78
88	Clone58	5.45	32.06	17.56	4.79	27.31	23.74	65.20	37.67	73.61	6.10	33.45	16.54
89	Clone93	3.65	38.12	16.88	3.43	23.36	18.07	75.30	27.61	61.60	10.05	50.49	17.43
90	CLONE99	3.94	34.43	13.95	4.40	28.69	20.96	46.97	48.50	90.70	3.77	26.67	19.63
91	Clone26	2.84	30.58	17.27	4.08	28.94	18.50	49.08	47.45	90.14	3.97	27.13	16.53
92	Clone7	3.84	37.34	16.30	2.92	25.91	13.98	62.62	25.89	85.87	6.00	42.37	20.61
93	Clone33	5.31	38.05	12.70	4.27	28.42	21.09	46.44	45.52	88.00	5.21	27.45	22.93
94	Clone47	6.88	39.80	16.20	4.18	28.60	18.98	47.20	44.00	95.11	3.73	23.87	22.39
95	Clone45	3.17	34.56	11.90	3.81	21.58	19.46	60.57	36.06	55.66	6.40	60.41	18.90
96	Clone77	4.09	34.88	15.25	3.75	26.88	17.26	51.07	40.72	87.12	5.84	30.44	22.90
97	Clone80	8.06	41.75	16.97	3.71	29.39	16.91	47.30	38.04	99.07	3.94	21.83	27.87
98	Clone35	5.51	38.84	14.79	3.59	28.24	17.46	60.09	28.88	72.09	4.49	40.87	23.68
99	Clone23	1.62	33.78	13.75	3.38	20.91	17.11	65.60	24.50	64.00	8.94	50.63	18.67
100	Clone24	4.49	36.19	11.65	4.80	25.01	23.85	48.54	56.53	65.57	5.38	25.82	18.74
101	Clone15	3.62	38.48	13.92	3.35	25.27	14.63	56.78	28.17	72.89	7.47	43.68	23.96
102	CLONE104	3.42	42.21	11.14	3.01	27.11	12.54	36.89	39.82	92.41	4.46	31.46	22.54
103	Clone74	3.01	35.09	14.79	4.38	25.60	20.41	43.55	46.54	76.02	3.70	28.74	30.72
104	Clone11	2.60	35.81	13.10	3.51	25.89	16.05	40.65	47.84	84.20	3.76	24.58	32.13
105	CLONE96	3.39	31.04	12.09	4.47	24.38	22.31	47.07	39.07	70.79	5.00	26.89	14.02
106	CLONE106	10.88	16.01	19.58	6.66	20.09	24.25	62.81	18.18	60.61	4.21	9.40	13.02

Table S6 4: Mean biochemical traits for Kericho. The values colored green indicates not significantly ($p<0.05$) different from each other for each trait.

	clone	Theobromine	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
1	CLONE101	4.34	34.42	19.97	4.90	29.91	26.00	82.72	50.63	91.58	10.72	43.63	11.96
2	Clone61	6.18	41.51	17.31	3.03	28.13	15.95	77.92	26.91	95.49	11.65	70.13	11.63
3	CLONE97	3.60	37.94	24.09	4.98	24.43	25.69	113.05	19.82	72.11	10.63	59.18	21.08
4	Clone69	3.11	39.98	20.10	5.34	29.66	27.44	86.75	42.21	89.00	6.68	39.15	20.14
5	Clone37	2.69	36.85	17.24	4.46	33.93	22.11	61.46	56.67	107.51	5.38	31.68	8.54
6	Clone55	5.89	44.08	16.94	4.27	37.00	19.09	58.59	38.93	128.60	4.14	32.02	15.03
7	Clone63	6.25	35.29	15.08	4.35	32.84	20.82	61.76	50.57	105.24	4.70	38.69	12.26
8	Clone68	5.52	39.09	18.02	3.46	33.66	18.13	75.50	35.08	99.29	5.03	45.67	17.59
9	Clone62	3.63	39.36	14.78	4.33	35.77	22.03	55.77	57.60	110.31	3.35	27.08	16.89
10	Clone57	5.35	40.59	14.47	4.95	28.60	25.84	62.85	59.97	90.53	5.52	38.55	21.05
11	CLONE95	5.50	39.83	17.73	4.44	32.97	21.35	62.45	48.03	102.35	5.33	36.93	19.43
12	Clone89	4.33	38.43	15.36	4.00	32.90	20.47	57.01	53.15	101.60	7.37	36.89	20.67
13	Clone29	3.54	35.89	14.12	4.69	33.50	22.47	52.41	63.83	101.52	3.03	32.58	15.15
14	Clone33	2.91	37.95	13.60	4.18	30.38	22.70	62.72	55.70	92.17	4.09	41.48	10.19
15	Clone70	3.20	35.22	23.68	7.73	30.92	44.25	93.66	39.16	62.88	5.22	32.69	14.75
16	Clone60	5.02	34.77	16.09	4.69	29.00	22.91	59.95	48.45	102.83	5.83	38.74	14.48
17	Clone51	4.61	39.92	17.71	4.56	34.72	23.57	63.76	45.64	96.28	7.26	35.63	17.89
18	Clone87	3.65	43.26	15.04	4.18	30.06	22.30	67.17	51.48	87.13	3.83	44.56	22.85
19	Clone9	4.34	35.27	17.24	3.77	24.24	20.75	78.73	37.11	74.82	11.10	58.94	18.38
20	Clone64	5.55	43.02	14.34	3.56	30.78	18.56	63.15	42.31	89.06	8.70	46.96	14.27
21	Clone18	2.93	38.69	15.47	4.10	31.92	22.08	58.49	59.96	92.01	5.28	28.65	14.77
22	Clone91	2.99	33.61	18.64	4.28	32.52	23.22	70.09	41.47	92.39	5.63	33.00	21.40
23	Clone53	2.86	30.09	22.31	4.71	27.02	24.30	99.38	21.61	65.12	8.66	51.61	19.71
24	Clone3	2.93	36.72	13.74	3.24	31.91	16.91	53.24	54.09	103.44	6.28	31.71	14.06
25	Clone72	3.14	28.96	15.80	5.11	30.84	26.60	55.86	65.20	89.72	4.67	24.49	8.98
26	Clone25	1.53	30.75	14.29	4.91	26.52	26.41	62.14	64.62	73.41	4.47	39.41	13.89
27	Clone46	4.28	42.14	13.89	4.09	29.51	21.65	58.66	54.49	90.28	5.24	36.31	23.00
28	Clone65	2.24	31.24	13.21	3.62	31.74	20.15	47.87	63.67	97.43	5.27	29.61	8.42
29	Clone82	2.93	38.11	16.12	4.42	32.97	23.47	58.81	53.29	94.40	3.66	26.25	21.21
30	Clone7	4.59	34.82	17.84	3.05	26.15	15.55	71.51	27.28	82.97	12.38	56.74	17.93
31	Clone67	4.37	39.50	13.22	3.57	25.22	16.63	52.40	46.38	98.00	6.58	47.33	21.75
32	Clone44	4.43	33.89	16.79	3.49	28.96	20.27	63.22	46.98	88.60	7.22	36.89	26.68
33	Clone35	4.17	41.64	16.11	3.74	28.90	17.20	65.63	38.95	86.23	7.84	46.34	12.37
34	Clone31	4.85	32.31	16.91	3.01	25.60	16.38	74.54	32.43	72.78	8.61	60.50	16.13
35	Clone49	2.80	43.62	15.15	4.20	32.86	22.01	59.61	45.07	100.83	6.75	22.85	18.93

36	CLONE100	3.11	37.00	13.50	3.29	27.64	16.12	52.96	46.37	99.80	3.74	42.52	15.38
37	Clone16	2.94	43.99	16.12	3.84	27.71	19.31	73.18	28.86	84.70	6.95	48.14	20.95
38	Clone21	3.39	37.83	14.79	3.61	23.44	19.38	69.63	36.07	71.63	10.86	57.52	21.21
39	Clone71	3.48	39.83	14.54	3.77	27.42	19.72	50.88	50.65	101.57	3.32	33.57	14.02
40	Clone85	3.44	34.98	13.70	3.15	32.11	16.12	39.96	60.81	102.29	4.38	29.95	21.04
41	Clone81	3.97	38.19	13.31	2.92	23.14	17.02	65.60	40.80	66.20	10.11	60.91	29.28
42	Clone17	3.09	37.85	13.85	3.23	23.55	18.10	65.47	38.99	67.24	14.38	55.94	22.74
43	Clone5	2.76	38.01	16.38	3.25	22.72	16.35	77.33	25.74	70.58	8.62	62.18	14.74
44	Clone48	4.76	46.25	15.57	3.52	32.68	18.27	50.28	41.83	109.65	2.87	27.81	15.47
45	Clone93	5.35	41.22	14.02	4.40	31.55	23.44	48.76	65.36	84.11	3.97	26.04	19.44
46	Clone94	2.62	34.70	15.20	4.39	26.60	23.29	73.72	33.05	70.17	5.65	50.44	12.92
47	Clone23	2.79	35.57	17.21	3.99	24.76	20.58	78.76	25.33	71.09	10.33	51.63	22.95
48	Clone19	1.92	36.48	12.50	3.53	27.72	17.85	56.30	52.07	81.70	6.02	40.61	11.98
49	Clone36	3.17	34.81	15.20	4.03	26.71	20.56	72.25	33.83	74.05	8.25	46.36	16.51
50	CLONE103	2.41	33.36	11.57	3.76	29.92	18.86	44.74	60.44	87.27	4.38	35.15	20.56
51	Clone34	2.97	36.28	17.46	4.00	22.40	18.83	66.55	33.45	75.95	9.08	54.16	16.19
52	Clone32	2.84	40.60	14.61	3.81	30.03	17.83	50.44	51.12	90.12	6.09	34.01	14.39
53	Clone45	2.76	32.14	11.37	3.86	24.17	19.46	57.67	42.00	64.13	11.82	59.94	15.02
54	Clone8	3.76	36.61	16.37	2.97	28.61	15.35	63.89	28.02	90.19	5.95	45.99	15.33
55	Clone66	3.93	36.72	13.41	4.06	31.91	20.64	48.70	55.71	86.94	6.40	26.89	12.22
56	Clone83	5.36	34.46	14.72	3.88	29.17	19.64	54.04	51.43	87.27	4.16	31.43	22.21
57	Clone1	4.82	38.05	15.86	3.98	30.58	17.11	52.00	47.45	91.56	5.24	32.74	21.25
58	Clone30	3.90	36.84	12.29	3.19	27.21	15.19	48.57	49.97	88.36	5.67	41.23	10.90
59	Clone86	5.40	43.21	15.82	4.57	28.55	24.03	61.30	46.77	77.00	6.87	31.57	29.51
60	Clone4	4.61	42.61	16.00	3.27	25.09	15.83	62.05	29.24	78.70	12.61	52.32	18.38
61	Clone28	4.98	42.39	16.40	3.72	26.89	19.65	74.83	25.15	72.54	7.69	48.96	21.21
62	CLONE102	6.12	40.12	15.09	4.12	26.87	20.64	52.09	44.94	96.09	3.42	31.61	18.29
63	Clone14	2.20	36.86	11.81	3.51	27.17	18.11	47.05	60.11	87.81	3.94	31.20	15.46
64	Clone42	5.62	39.62	11.61	3.82	26.59	19.35	43.21	59.21	88.05	6.61	31.20	21.10
65	CLONE99	3.32	34.25	12.52	4.09	26.67	20.61	49.45	56.02	82.85	4.92	33.17	14.62
66	Clone58	3.60	34.39	14.98	4.38	27.88	22.91	58.10	45.89	81.37	6.52	30.50	23.46
67	Clone13	3.67	37.11	12.20	3.78	28.40	16.57	43.31	56.30	84.86	7.33	36.21	15.24
68	Clone22	4.25	41.64	15.17	2.74	26.91	13.45	57.52	27.13	91.18	8.84	47.40	19.85
69	Clone92	5.16	39.60	16.09	4.67	28.67	25.28	58.68	47.02	81.66	4.35	26.63	16.24
70	Clone2	3.37	33.17	18.56	3.35	20.84	17.19	82.45	22.30	62.23	10.76	56.29	12.26
71	Clone12	4.11	39.11	18.28	4.93	29.32	23.34	55.52	42.84	88.65	3.26	25.55	20.54
72	CLONE98	4.16	38.42	15.15	3.27	22.77	17.13	71.31	25.67	71.94	6.97	52.60	17.51
73	Clone75	3.12	36.29	11.15	2.88	26.35	15.38	46.45	48.85	81.64	7.95	40.96	22.40
74	Clone76	2.84	33.52	14.92	4.25	29.54	23.11	58.58	47.14	73.61	3.64	31.80	15.02
75	Clone40	3.89	38.55	17.80	3.55	27.09	19.14	68.17	28.58	78.04	7.86	38.40	17.50
76	Clone78	4.37	35.71	13.04	3.57	26.39	20.11	49.85	56.86	77.53	5.88	29.94	27.44
77	Clone50	4.72	35.99	16.78	4.00	28.25	20.87	55.71	45.73	84.96	3.46	26.75	16.95
78	Clone90	3.79	40.14	13.62	3.52	30.42	18.23	45.86	46.86	88.39	5.40	28.54	18.16

79	Clone54	5.03	37.26	14.56	3.42	27.16	17.85	54.59	36.35	88.51	6.29	32.71	19.06
80	Clone20	3.25	33.93	13.29	5.82	27.86	28.32	54.05	51.30	62.73	4.83	33.40	18.08
81	Clone24	2.47	35.03	11.28	4.64	25.01	24.39	48.75	58.54	72.28	4.28	28.16	15.85
82	Clone38	4.15	35.95	12.00	3.19	27.13	16.11	47.69	44.08	87.32	4.46	32.30	25.41
83	Clone84	4.70	36.33	12.89	3.55	31.65	17.65	43.83	47.38	89.15	3.54	25.49	27.96
84	Clone79	2.50	30.26	13.13	3.79	26.06	18.40	48.31	49.33	81.05	4.25	31.03	17.97
85	Clone52	3.05	41.47	12.21	3.87	27.65	17.43	42.10	56.01	80.97	5.18	27.58	18.24
86	Clone26	2.04	31.09	13.68	3.51	25.94	17.33	45.65	47.66	82.65	4.35	33.27	13.56
87	Clone77	3.62	35.59	12.80	2.83	27.72	13.82	44.59	44.54	86.11	5.98	32.86	12.99
88	Clone73	3.09	36.26	12.14	3.87	26.95	18.77	43.24	54.32	82.11	3.60	26.31	24.56
89	Clone56	3.85	31.90	15.46	2.73	22.84	14.00	73.41	14.28	76.19	7.94	45.51	20.43
90	Clone10	2.79	27.00	13.68	3.40	26.47	17.64	46.02	47.11	85.12	3.30	27.96	13.28
91	Clone27	2.71	39.31	11.34	3.49	27.33	17.08	40.34	52.93	91.49	2.34	20.58	18.03
92	Clone74	4.66	33.20	15.28	4.32	27.91	21.49	47.39	57.01	69.43	4.55	24.27	21.73
93	Clone88	3.79	36.50	12.00	4.14	27.65	21.70	42.27	61.09	67.57	4.19	27.39	17.75
94	Clone59	5.57	39.84	16.83	3.54	26.54	17.94	55.32	32.52	83.53	4.68	30.15	21.72
95	Clone47	4.36	40.59	14.72	3.38	27.53	17.96	50.86	37.31	80.73	3.01	31.73	22.24
96	Clone39	2.55	33.07	16.18	2.83	24.14	12.69	61.26	17.92	83.69	8.24	39.70	11.79
97	Clone43	3.82	32.72	14.56	2.89	25.38	14.21	56.51	27.59	75.74	7.68	40.16	26.42
98	Clone80	4.95	39.80	14.61	3.39	27.93	17.12	43.49	38.57	89.19	4.64	23.56	26.34
99	Clone11	2.22	34.42	11.37	3.17	24.67	16.11	40.38	51.82	76.34	4.43	29.67	25.57
100	Clone6	1.62	31.52	10.55	3.08	21.74	16.98	43.58	49.57	72.01	3.90	33.66	14.39
101	Clone41	2.71	37.21	16.01	2.80	25.15	14.67	56.83	22.39	70.26	4.40	36.63	21.52
102	CLONE96	2.30	29.71	11.26	4.07	24.05	21.77	46.41	44.11	62.22	4.77	26.81	13.38
103	Clone15	3.26	36.74	11.85	2.73	19.75	12.66	51.04	25.74	61.09	6.74	45.26	19.45

Genomic prediction and genome wide association mapping for quality in tea (*Camellia sinensis* (L.) O. Kuntze) using genotyping-by-sequencing

Lubanga M. Nelson[b,c*], De-Silva Jacquie[d], Massawe Festo[b], Mayes Sean[a]

*nelsonlubanga@gmail.com

[a]School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK.

[b]School of biosciences, University of Nottingham, Malaysia Campus Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[c]Unilever Tea Kenya Ltd., Nakuru Road, Kericho, 20-20200, Kenya.

[d]Unilever Research & Development, Colworth Science Park, Sharnbrook, Bedford MK441LQ, UK.

Abstract

Genetic improvement of tea quality is an important objective in tea breeding. However, the perennial nature of tea and the subjectivity of the sensory evaluation method limit selection to few breeding lines. Marker assisted selection can improve the selection efficiency, especially for traits that are difficult, challenging and costly to phenotype such as tea quality. Here, we used genome-wide association study (GWAS) to identify significant SNPs and candidate genes associated with tea quality and explored the potential application of genomic selection (GS) in a tea breeding programme. Our analyses identified 64 SNP markers associated with quality related traits in tea, and 44 SNP markers with R^2>0.2 were further selected for identification of candidate genes. The potential candidate genes included several transferases, cytochrome P450 704C1-like proteins, E3 ubiquitin protein ligases, ATP-dependent zinc metalloprotease and exopolygalacturonases. Based on gene ontology annotation, 217 sequences were annotated, while 311 sequences had blast hits. Gene ontology terms for the 217 annotated marker sequences were assigned to biological processes (38%), cellular component (34%) and molecular functions (28%). In total, 65 KEGG

pathways and 6 major enzyme groups; hydrolases (45%), transferases (27%), oxidoreductases (14%), ligases (5%), lyases (4%) and isomerases (2%) were identified. GS prediction accuracies of Ridge regression best linear unbiased predictor (RRBLUP), Bayesian least absolute shrinkage and selection operator (BayesLASSO), BayesA, BayesB and BayesCπ were compared. Repeatability was high for all the traits except catechin. The traits with the highest mean prediction accuracies across the models were; Epigallocatechin gallate (ECGG) (0.62), Theanine (0.62), Epicatechin (EC) (0.61), Epicatechin gallate (ECG) (0.61) and theobromine (0.61), while the traits with the lowest prediction accuracies were Gallocatechin (GC) (0.33), catechin (0.36) and Gallic acid (GA) (0.42). The performance of all the GS models were almost the same, with RRBLUP (0.53), BayesLASSO (0.53) and BayesA (0.53) performing slightly better than BayesCπ (0.52) and BayesCπ (0.51). However, RRBLUP is recommended because it is computationally simple to use compared to the other models. Our results show the potential for tea quality to be predicted more efficiently using GS. The candidate genes and the associated SNPs provide valuable resources for future studies to breed high quality tea varieties and to understand the basis of tea quality on a chemical level, hence complementing the current organoleptic method of tea-tasting. The identified SNP markers could be further fine mapped to evaluate their potential involvement in tea quality.

7.1. Introduction

Tea (*Camellia sinensis* (L.) O. Kuntze) quality is an important attribute and is defined by the flavour and colour of tea liquor (hue) along with appearance of dry tea (leaf) (Zheng et al., 2016a). Flavour comprises of taste, mouthfeel and aroma (Lawless et al., 2010). Taste of tea is characterized by the astringency, bitterness, mellowness and slight sweetness (Lee et al., 2007). Mouth-feel is the heaviness, thickness and strength of tea liquor, while aroma is influenced by more than 600 volatile compounds known to be present in tea (Zheng et al., 2016a). Taste, mouthfeel, colour and aroma are generated by biochemical compounds present in fresh tea shoots such as flavonoids, alkaloids, amino acids and volatile compounds (Chen et al., 2018; Borse, 2012).

Gene expression, functional annotation and proteins associated with quality traits in tea (*Camellia sinensis* (L.) O. Kuntze) have been studied, based on a knowledge of which biochemical pathways are likely to be involved and some candidate genes have been partially identified using linkage mapping studies (Xie et al., 2019; Koech et al., 2019; Koech et al., 2018b). For instance, Koech et al. (2019) studied the functional annotation of putative QTLs associated with black tea quality and drought tolerance in tea (*Camellia sinensis* (L.) O. Kuntze) using a linkage map derived from 1,421 DArTseq markers. They identified 47 putative QTLs comprising of 6 caffeine, 25 catechins, 3 theaflavins, 9 QTLs for tea taster score and 3 QTLs for percent relative water contents were detected. Additionally, they found 84 unigenes and a few putative proteins that could be involved in flavonoid and alkaloid biosynthesis. The QTLs associated with caffeine, individual catechins and theaflavins were clustered in the LG02 and LG04 map regions (Koech et al., 2018b). However, these findings are less sensitive due to the genetic limitations imposed by using a bi-parental mapping population that were derived from a linkage map. The mapping populations comprised of 261 F_1 clonal progenies from two segregating populations comprising of TRFK St 504 and TRFK St 524 (Koech et al., 2018b). Bi-parental mapping populations used in QTL mapping capture a small portion of the total genetic variation in a breeding programme and also produces low resolution locational information because only a single generation of recombination is evaluated (Crossa et al., 2017).

GWAS is a powerful method used to detect and map QTLs based on significant association of markers with phenotypic traits (Gupta et al., 2005). It exploits historical linkage disequilibrium to identify marker-trait relationships within a natural or breeding population (Pasam et al., 2012), resulting in higher mapping resolution (Yu et al., 2011; Stich et al., 2010). Important genes influencing tea quality can be dissected using GWAS (Jin et al., 2016). For instance, the SNP marker SNP4318 was identified to be significantly associated with caffeine content in tea and its' function validated by site-directed mutagenesis (Jin et al., 2016). However, although this is the only association mapping studies reported in tea, only a limited set of genotypes (44 accessions) were studied.

GS is a modern breeding approach whereby a model based on genome-wide markers are used to estimate individual breeding values of untested genotypes (Lorenz et al., 2011a; Jannink et al., 2010b;

Heffner et al., 2009b; Meuwissen et al., 2001). GEBVs of the next generation of untested genotypes with only genotypic information are computed using the constructed model and these are used for selection of superior individuals without direct phenotypic evaluation (Meuwissen et al., 2001). In GS, the number of markers are usually greater than the number of phenotypic measurements of the traits of interest, hence there are more predictor variables compared to phenotypes, hence creating a "large p and small n problem" (Lorenz et al., 2011a). Statistical models that have been developed to solve the problem of having large numbers of molecular markers and fewer phenotypes include ridge regression best linear unbiased predictor (rrBLUP), genomic best linear unbiased predictor (G-BLUP), the Bayesian models (BayesA, BayesB, BayesC, BayesLASSO) and machine learning (Wang et al., 2018b; Lorenz et al., 2011a). GS models have successfully been developed for predicting traits for many crops (Müller et al., 2019; Grattapaglia et al., 2018; Cerrudo et al., 2018; Tan et al., 2017; Sverrisdóttir et al., 2017; Juliana et al., 2017; Bassi et al., 2016; El-Dien et al., 2015; Resende et al., 2012a). GS can potentially reduce the length of the tea breeding cycle in tea and increase gains per unit time through early selection, with the GS model being used to carry out 1-2 rounds of selection based on genotype alone, before the need to rebuild the model due to the change in allelic frequencies caused by selection.

Our study was intended to accomplish the following objectives: 1) perform GWAS to identify potential SNPs associated with biochemical traits in tea, 2) identify potential candidate genes associated with tea quality for future studies, 3) assess the effectiveness and potential of GBS as a genotyping platform for GS studies in tea, 4) estimate genomic and broad sense heritabilities of twelve quality traits assessed by Nuclear Magnetic Resonance spectroscopy (NMR) in tea, 5) compare the prediction accuracies of five GS models: Ridge regression best linear unbiased predictor (RRBLUP), Bayesian least absolute shrinkage and selection operator (BayesLASSO), BayesA, BayesB and BayesCπ .

7.2. Materials and methods

7.2.1. Plant materials and phenotyping

Genotypes used in this study consisted of 103 tea varieties (clones), present in the UTK breeding programme clonal field trials (CFTs) at Kericho (0^0 22' S and 35^0 17' E), which is located at 2005 meters above sea level and replicated at Jamji (0^0 28' S and 35^0 11'E), situated at 1733 meters above sea level. The training population was then phenotyped using NMR spectroscopy for 12 biochemical traits namely; theobromine, caffeine, theogallin, gallic acid (GA), epicatechin (EC), gallocatechin gallate (GCG), epicatechin gallate (ECG), epigallocatechin gallate (EGCG), epigallocatechin (EGC), theanine, catechin (C) and gallocatechin (GC) according to Le Gall et al. (2004). The mean values of the phenotypic data used in this study are presented in (Table S6. 2). For each of the trait, best linear unbiased predictors (BLUPs) using their replicated data at each site were generated using linear mixed models in R (R. Core, 2015). The restricted maximum likelihood (REML) method was used to estimate variance components assuming a random effect model. BLUP values were estimated for each trait, by treating genotype and site as a random effect. The Pearson's correlation coefficient (r) was calculated between all the twelve traits. Shapiro-Wilk test for normally distributed data was conducted for all traits in R (R. Core, 2015). All traits except theanine, EGCG, caffeine and EC were statistically significant with p-values under 0.05, and were normalized using Log base 10 in R (R. Core, 2015).

7.2.2. Genotyping

GBS was used to genotype all the 103 genotypes in the training population and was conducted at the Cornell University Institute of Genomic Diversity. Green leaf samples were collected early in the morning from the CFTs, freeze-dried for 3 days and stored in waterproof aluminum sachets. The freeze-dried samples were then shipped to ADNid laboratories in France for DNA extraction and quantification using the DNeasy 96 Plant Kit (QIAGEN). High-quality DNA was then sent to Cornell University's Institute of Genomic Diversity for genotyping using GBS. A multiplexed, high-throughput GBS procedure

was conducted according to the procedure of Elshire et al. (2011). Sequence data were obtained from 96-plex Illumina HiSeq2000 runs. For genomic complexity reduction, the PstI restriction endonuclease was used. The SNP discovery and genotype calling was performed using the TASSEL (version 5.2.48)-GBS pipeline which utilizes the Universal Network Enabled Analysis Kit (UNEAK) SNP calling algorithm, tailored to species with no reference genome (Lu et al., 2013). Individuals with not more than 19% missing SNPs were selected and missing SNPs were imputed using kNN algorithms in R using VIM software (R. Core, 2015). The SNP markers were then recoded as -1, 0 and 1, corresponding to homozygous minor alleles, heterozygous and homozygous major alleles, respectively. A total of 2779 SNPs from the 103-tea genotypes were used in the present study.

7.2.3. Homology search, functional annotation and KEGG pathway assignment

All the 2779 SNP marker sequence tags were subjected to BLASTN search against the published draft tea genomes (Xia et al., 2019) using CLC Genomics Workbench version 7. Functional annotation of all the 2779 SNP markers was performed using BLASTX against the non-redundant GenBank protein sequence database with a threshold E-value of 10^{-6} using Omicsbox 1.0.34 software (Conesa et al., 2008; Conesa et al., 2005). The procedure used for BLASTX was; functional annotations, GO mapping, annotation, InterProScan, Merge InterProScan with GO mapping, Enzyme Code Mapping and KEGG Maps Functional Analysis. This has provided information on their potential biological process, molecular function and cellular component. Additionally, the KEGG pathway analysis identified metabolic pathways involved in biosynthesis of secondary metabolites and proteins in tea.

7.2.4. Population structure

All the statistical analysis was done in R (R. Core, 2015). To visualize the relatedness and population structure among the 103 genotypes, the realized genomic relationship matrix was created from the genotype matrix using the A. mat function in R via the rrBLUP (Endelman, 2011). Principle component

analysis (PCA) was determined using the 2779 SNP markers and was estimated using the k-means clustering function in R and the first two principle components were plotted (R. Core, 2015).

7.2.5. Heritability

Heritability was estimated from linear mixed models. Broad sense heritability was calculated as $h^2 = \sigma^2a/ (\sigma^2a+\sigma^2sg+\sigma^2e)$ (El-Dien et al., 2015; Bekele et al., 2014). Variance components were estimated using the restricted maximum likelihood method (REML) using the lmer package in R (R. Core, 2015). Where σ^2a is the genetic variance, σ^2sg is the variance due to genotype by environment interaction and σ^2e is the error variance.

Genomic heritability was estimated as the ratio of the genomic and the phenotypic variance (VanRaden, 2008) and was estimated using a genomic relationship matrix (de los Campos et al., 2015; de los Campos et al., 2012).

7.2.6. Estimates of repeatability of traits

To estimate the repeatability of the NMR spectroscopy, 3 technical replicates of the 12 biochemical traits from 31 samples were analyzed (Table S7. 2.). The data was subjected to linear mixed effects models to estimate variance components. From the analysis of variance, repeatability of each trait was estimated using the following formula as used by de Villemereuil et al. (2018).

Repeatability = $\sigma^2a/ (\sigma^2a+\sigma^2e)$

Where σ^2a is the genotypic variance while σ^2e is the residual variance.

7.2.7. Genome wide association mapping and candidate gene selection

To identify the marker trait association (MTA) for all the biochemical compounds, GWAS was performed using the 2779 SNP markers and the 12 biochemical traits using the mixed linear model (MLM) procedure (Yu et al., 2006a), accounting for both population structure and kinship in TASSEL (Trait

Analysis by aSSociation Evolution and Linkage) version 5.2.48 (Bradbury et al., 2007). Since the population structure can result in false associations, it was factored in the analysis by using the first five principal components (Price et al., 2006) as estimated using the default settings in TASSEL.

The 2779 SNPs were used for GWAS analysis, and the formula used to select significant markers was P = 1/total number of SNPs markers = 3.59E-04 (Zhu et al., 2018; Ma et al., 2016). GWAS was conducted using the mean biochemical data from samples at Kericho and Jamji sites (Table S7. 1). To identify candidate genes, the position of significant SNPs from the GWAS results with $R^2 > 0.2$ were mapped to the reference genome using CLC Genomics Workbench version 7 with default settings and the reference genome was downloaded from Tea Plant Information Archive database (http://tpia.teaplant.org) (Xia et al., 2019). For the GWAS analysis, the mapped SNP markers were with scaffolds since the reference genome is yet to be at pseudo-chromosomal level. The two available draft tea genomes have many scaffolds and a high proportion of sequence repeat regions of 64% (Wei et al., 2018b) and 80.9% (Xia et al., 2017b). However, the scaffold positions for each significant marker were identified and are reported in this study.

7.2.8. Prediction accuracies of the models for the biochemical traits

This study investigated the performance of 5 GS prediction models that differ on the assumptions of marker effects. The GS models selected were Ridge regression best linear unbiased predictor (RRBLUP)(Endelman, 2011; Meuwissen et al., 2001), the Bayesian least absolute shrinkage and selection operator (LASSO) (Usai et al., 2009), BayesA (Meuwissen et al., 2001) BayesB (Meuwissen et al., 2001) and BayesCπ (Meuwissen et al., 2001).

RRBLUP is computationally similar to genomic BLUP (GBLUP) and it assumes that marker effects are equally shrunk and normally distributed with the same variance (Meuwissen et al., 2001). It is an infinitesimal model and assumes that all the markers have small effects and have non-zero variance. On the other hand, the Bayesian models assume marker effects are independent, with different shrinkage and variance. BayesA and BayesLASSO assume that all markers have a non-zero effect, and the marker variances are derived from a scaled inverted chi-square and double-exponential distributions, respectively.

Both BayesB and BayesCπ are variable selection models since they are derived from two component mixtures with a point of mass at zero that can either be a scaled-t or a normal distributions, respectively (Habier et al., 2011).

All the GS models were implemented using the BGLR package with default settings for priors (de los Campos et al., 2016) in R version 3.5.3 (R. Core, 2015). The GS analysis in BGLR was set for 12000 iterations and a burn-in setting of 2000. The predictive accuracy of all the GS models was estimated using a 5-fold cross-validation approach for all the traits. The data was randomly divided into 5 subsections, and one subset was also used as a distinct validation set (corresponding to 20% of the genotypes), while the remaining four groups (80% of all the genotypes) were used as training population for fitting the GS models. This process was repeated, each time with another subset, until all subsets had been used in both training and validation steps. Each analysis was repeated with 10 different cross-validation groupings and the mean GEBVs for the 10 subsets was calculated. The accuracy of the GS models was estimated as the Pearson correlation between the mean GEBVs and the observed phenotypes (biochemical traits); *r(GEBV:y).*

7.3. Results

7.3.1. Biochemical traits

All the biochemical traits for all the 103 genotypes were analyzed using NMR spectroscopy. The mean (mg per gram) biochemical contents, coefficient of variation and ranges are presented in Table 15. The coefficients of variation ranged from 10.1% to 56.5%, signifying broad phenotypic variation. ANOVA revealed highly significant differences ($p<0.001$) among all the traits, signifying existence of genetic variation that can be exploited for breeding (Table S7. 3). When tested with Shapiro-Wilk, all traits except theanine, EGCG, caffeine and EC are statistically significant with p-values under 0.05, implying that they did not follow a normal distribution (Table 16). Once the data was normalized, all traits followed a normal distribution and were then used for GWAS analysis (Figure 17).

Table 15. Mean biochemical values (mg per gram), coefficient of variation (CV), and maximum and minimum values of the biochemical traits (mg per gram) across the sites.

Trait	CV %)	Mean (mg per gram)	Maximum (mg per gram)	Minimum (mg per gram)
Caffeine	10.1	37.83	46.66	27.92
Catechin	56.5	6.15	12.89	2.93
EC	19.6	28.81	34.95	22.51
ECG	13.7	60.65	112.61	40.51
EGC	18.9	43.42	69.30	17.13
EGCG	14.7	87.29	120.82	59.89
Gallic acid	22.8	4.08	5.82	2.94
GC	18.4	37.87	63.28	22.36
GCG	18.1	19.97	33.01	13.65
Theanine	29.9	19.09	29.82	10.70
Theobromine	38.7	4.28	6.84	2.10
Theogallin	18.5	15.83	24.39	11.47

Table 16. Shapiro-Wilk Test for normal distribution.

Trait	Shapiro-Wilk	P value
Catechin	0.9202***	0.00001
ECG	0.92832***	0.00003
GC	0.93653***	0.00009
GCG	0.95491***	0.00146
Theogallin	0.96214***	0.00484
EGC	0.96746**	0.01220
Theobromine	0.96761**	0.01252
GA	0.97224*	0.02886
Theanine	0.98	0.12200
Caffeine	0.99	0.34840
EGCG	0.99	0.49280
EC	0.99	0.75250

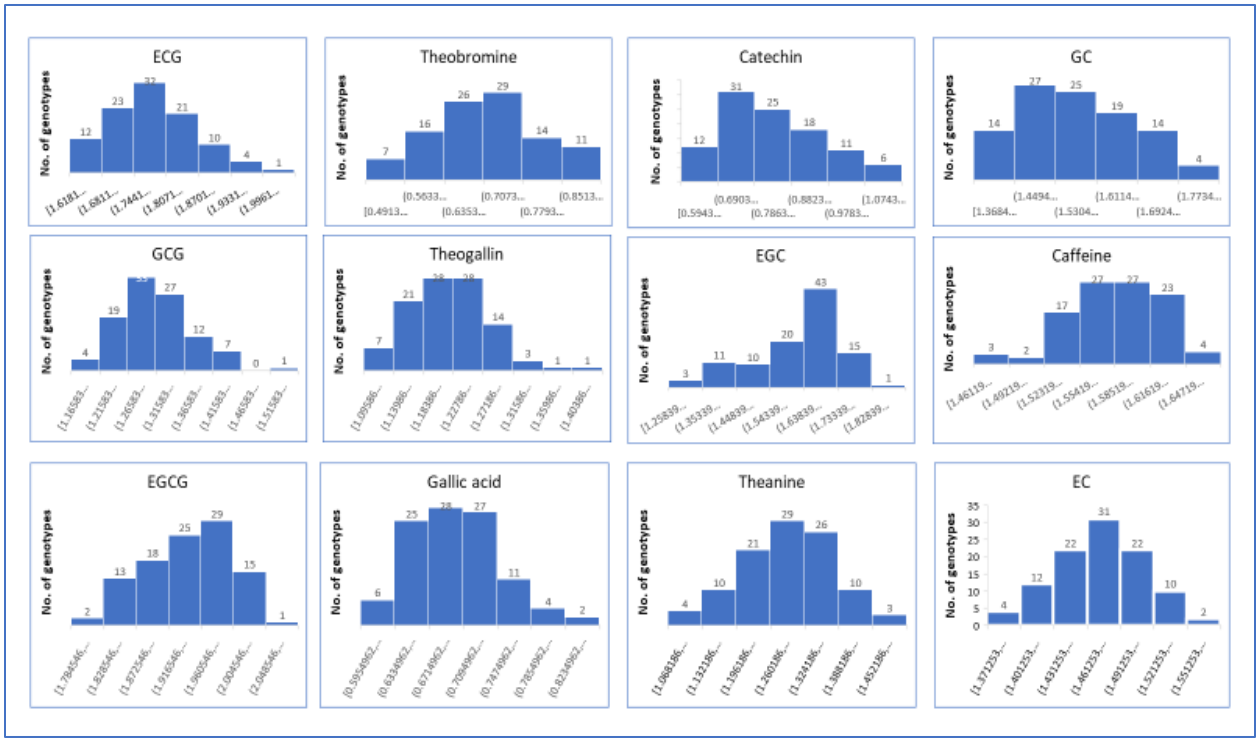

Figure 17. Biochemical levels histograms showing the phenotype distributions for the 103-training population. The biochemical contents were evaluated using NMR spectroscopy and measured in milligram per gram tissue.

7.3.2. Repeatability of the biochemical measurements

From the data collected using NMR spectroscopy for the 12 biochemical traits, the coefficients of repeatability (r) were determined (Table 17). Repeatability coefficient (r) measures the ability of individuals to repeatedly express the measured traits over several periods of time. In this study, repeatability was estimated to inform the individual consistency of the biochemical traits by the NMR process. It was observed that the repeatability coefficients for all the traits were high except catechin and theobromine, signifying that the NMR spectroscopy method used was reliable. Theanine and ECG had the highest repeatability estimate at 0.98 and 0.96 respectively, while catechin had the lowest repeatability estimates at 0.34 (Table 17).

Table 17. Repeatability coefficient of the traits.

Trait	Repeatability coefficients (r)
Theanine	0.98
ECG	0.96
EGC	0.95
EGCG	0.95
Theogallin	0.93
Caffeine	0.93
GCG	0.91
GC	0.9
GA	0.9
EC	0.9
Theobromine	0.75
Catechin	0.34

7.3.3. Genotypic data analysis

A total of 155 billion base pair of good barcoded raw DNA sequence data were generated in GBS, with an average of 2 million reads per genotype. TASSEL UNEAK SNP calling algorithms was used to determine SNP polymorphism, resulting in 82,254 SNPs. Nature Source Improved Plants (NSIP) applied an inhouse SNP calling algorithms to further filter to leave a high quality 2779 SNP dataset by decreasing error rate and increasing reliability (Professor Steve Tanksley, Pers. com, May 2016, NSIP). The average read count per sample per locus (SNP) was 39 and the average missing data is 19% per individual and per marker.

The genome of the two main varieties of tea; *Camellia sinensis* var. *sinensis* (CSS; Chinese type) (Wei et al., 2018b) and *Camellia sinensis* var. *assamica* (CSA; Assam type) (Xia et al., 2017c) have been sequenced and the draft reports indicate that they contains 64% and 80.9% sequence repeat regions with 14,051 and 37,618 scaffolds respectively. Among the 2779 sequence tags, only 929 SNPs mapped to each genome.

7.3.4. Assessment of the population structure

The population used in this study consists of tea genotypes with diverse quality properties. Known high-quality clones and poor-quality clones were selected. There were two clear population structures from the heat map of the realized genomic relationship matrix (Figure 18). This was also confirmed by the principal component analysis (PCA) of the genotype data, with the first two principal components explaining 30% and 11%, respectively of the total marker variation, making a total of 41% (Figure 19).

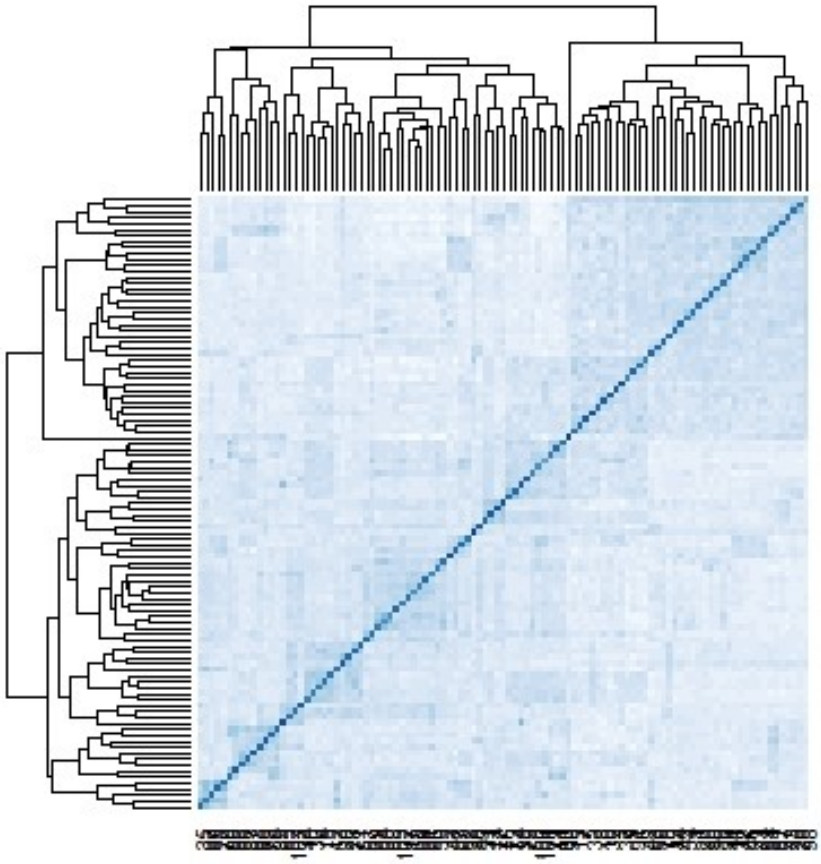

Figure 18. Heat map of the realized genomic relationship matrix of the 103 tea (*C. sinensis*) genotypes.

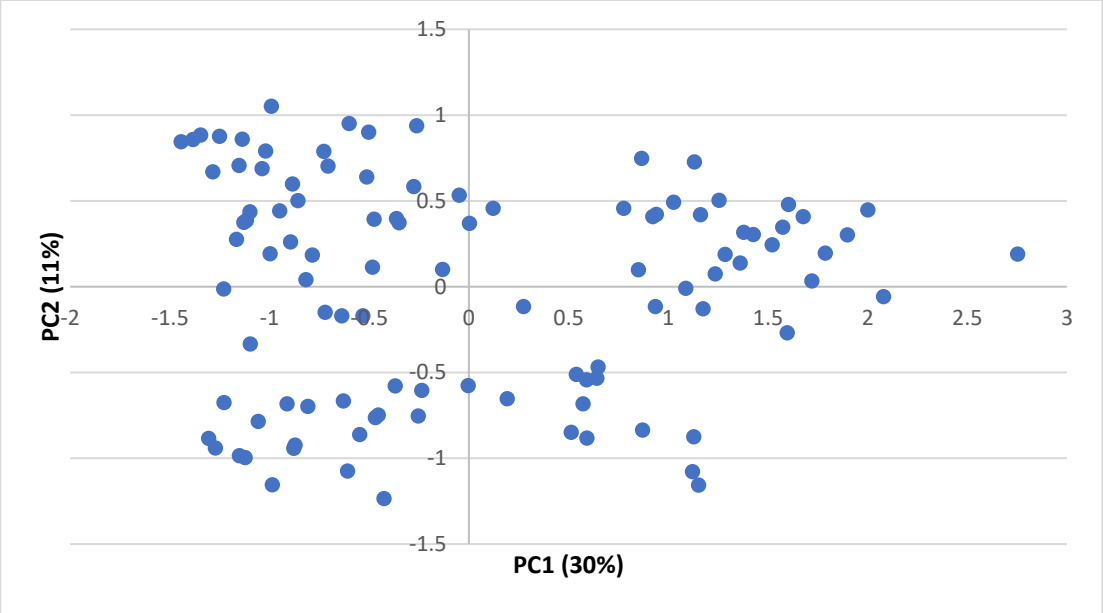

Figure 19. A PCA plot of the 2779 SNP markers for the 103 genotypes. The variance explained by principal components 1 and 2 are indicated in parentheses.

7.3.5. Sequence homologies, functional annotation and gene ontology mapping

Based on gene ontology annotation, 217 sequences (7.8% of the total sequences) were annotated, while 311 sequences (11.19% of the total sequences) had blast hits when blasted against the NCBI's nr database (Figure 20). The 217 annotated sequences that were assigned GO terms comprised of biological processes (38%), cellular component (34%) and molecular functions (28%) (Figure 21). The GO terms were analyzed from 43 different functional groups (level 3) (Figure 22). Within the biological process cluster, the dominant GO terms were organic substance metabolic process, primary metabolic process, cellular metabolic process, nitrogen compound metabolic process and biosynthetic process. Within the molecular function group, the largest GO terms were heterocyclic compound binding, organic cyclic compound binding, ion binding and hydrolase activity. For the cellular component, the largest GO terms were intrinsic component of membrane, intracellular, intracellular organelle, intracellular part and membrane-bounded organelle (Figure 22).

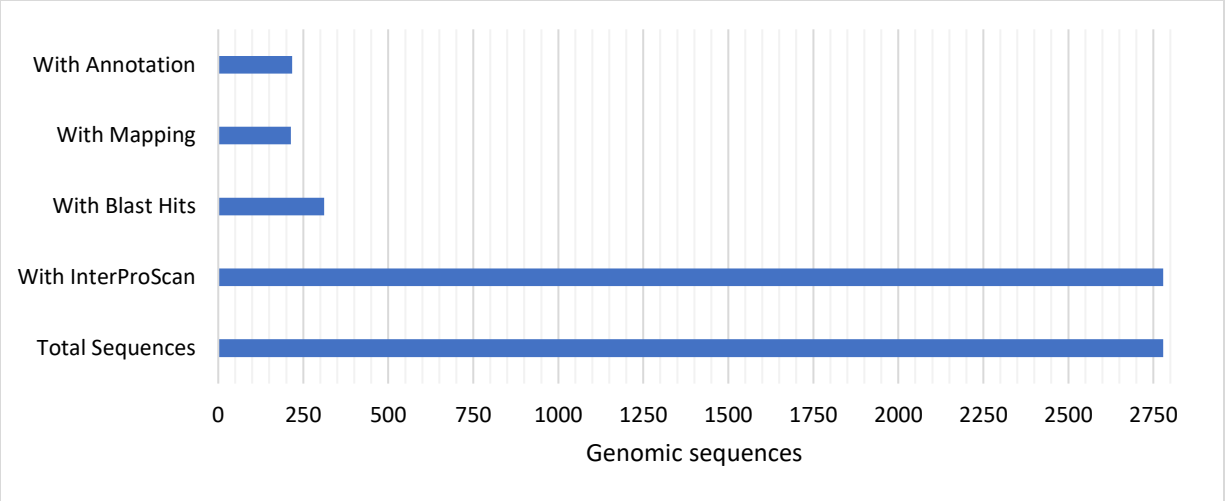

Figure 20. Analysis progress of the blasted 2779 GBS sequence tag against the NCBI's nucleotide database.

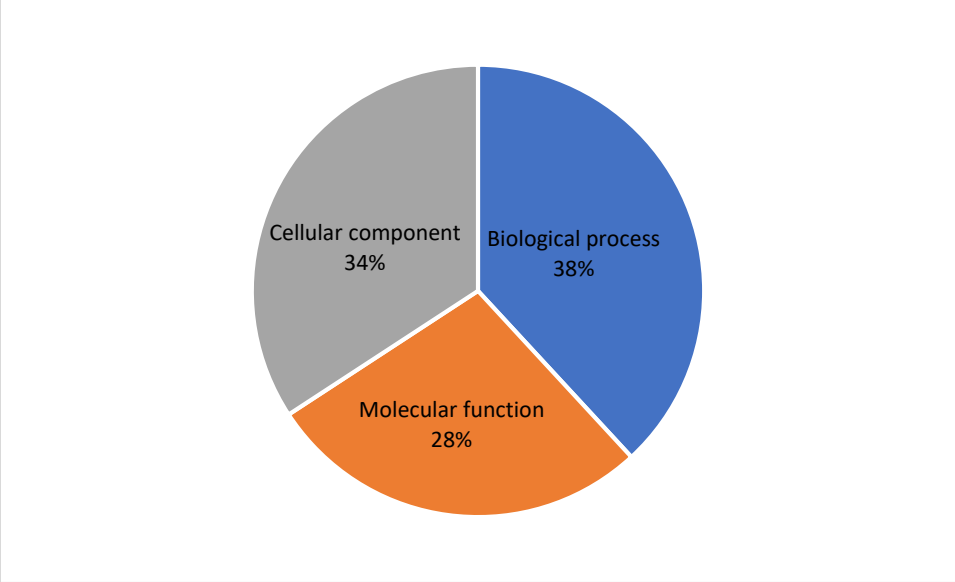

Figure 21. GO annotation of the 217 *C. Sinensis* sequences (level 3) that were grouped into biological process, molecular function and cellular component.

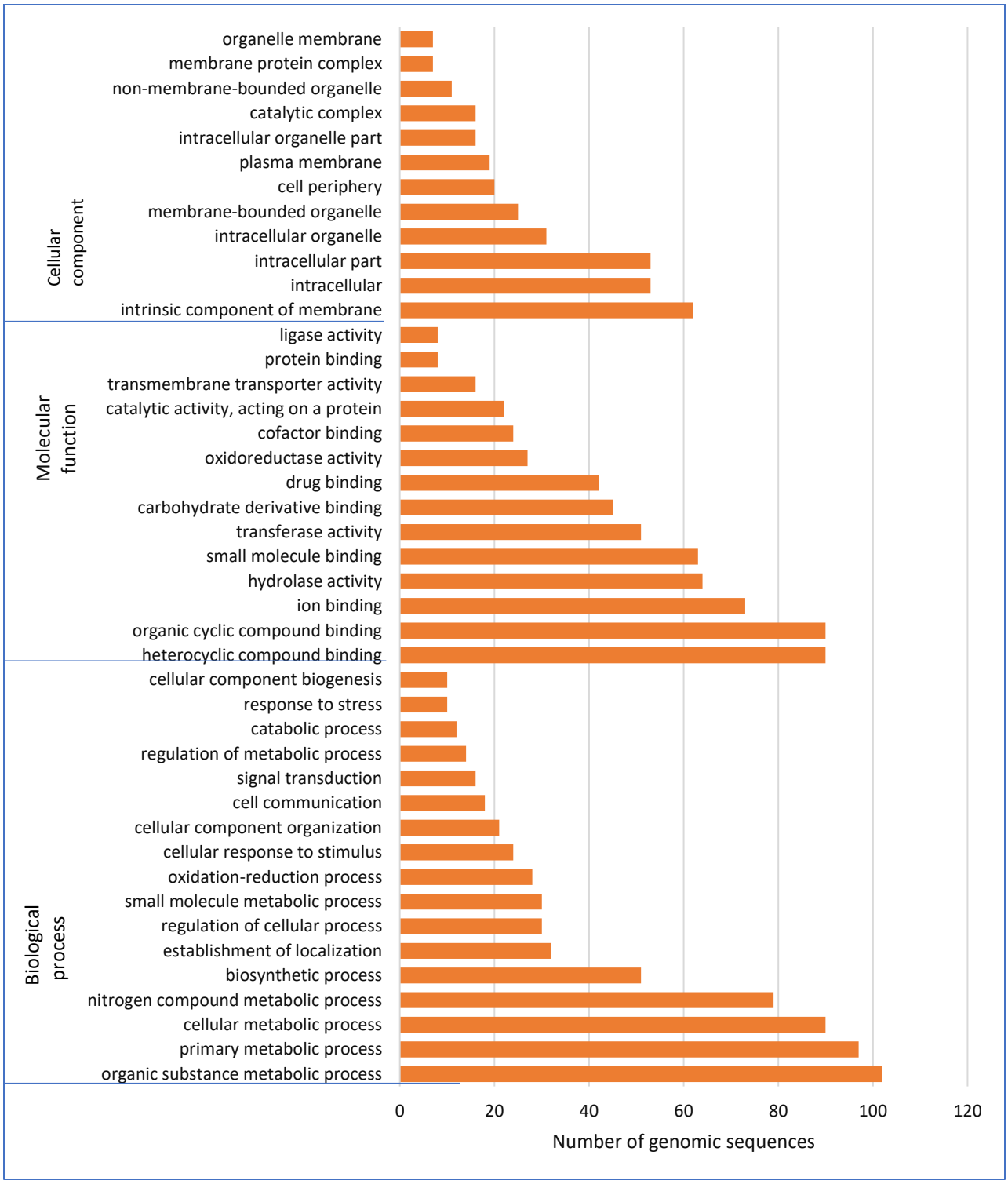

Figure 22. Summary of GO functional classification of the *Camellia sinensis* sequences based on high score (in level 3) in the three main categories: biological process, molecular function and cellular component.

7.3.6. KEGG pathway analysis

In order to try to understand the biological functions of the genes involved in metabolism of secondary metabolites in tea (*Camellia sinensis* (L.) O. Kuntze) the sequences were mapped to the KEGG pathways information in the Omicsbox software version 1.0.34. Generally, a large number of functional enzymes and regulatory proteins were identified (Figure 23; Table S7. 5 and Table S7. 6). In total, 217 *C. sinensis* sequences were assigned to 65 KEGG maps and 573 EC (Table S7. 5 and Table S7. 6).

The pathways with the largest representation were; biosynthesis of antibiotics, purine metabolism, drug metabolism, glutathione metabolism, fatty acid biosynthesis, pyruvate metabolism, arginine biosynthesis, carbon fixation pathways in prokaryotes, cysteine and methionine metabolism, biotin metabolism, pantothenate and CoA biosynthesis, propanoate metabolism and other pathways (Figure 23; Table S7.5 and Table S7.6).

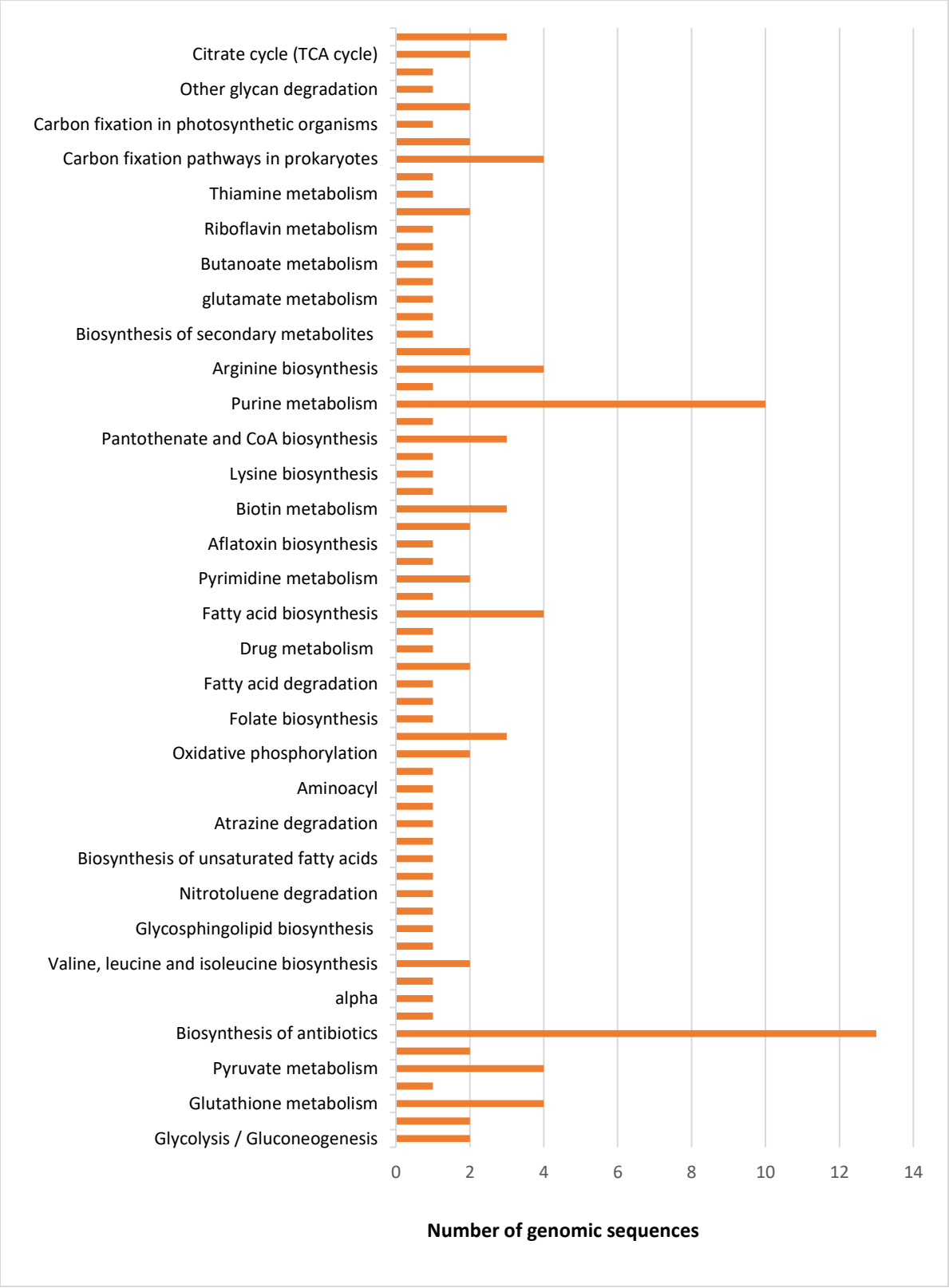

Figure 23. KEGG pathway distribution.

Six major enzymes were identified, and the most abundant were hydrolases (45%), followed by transferases (27%) and oxidoreductases (14%). The other less abundant enzymes were ligases (5%), lyases (4%), and isomerases (2%) (Figure 24; Table S7. 6).

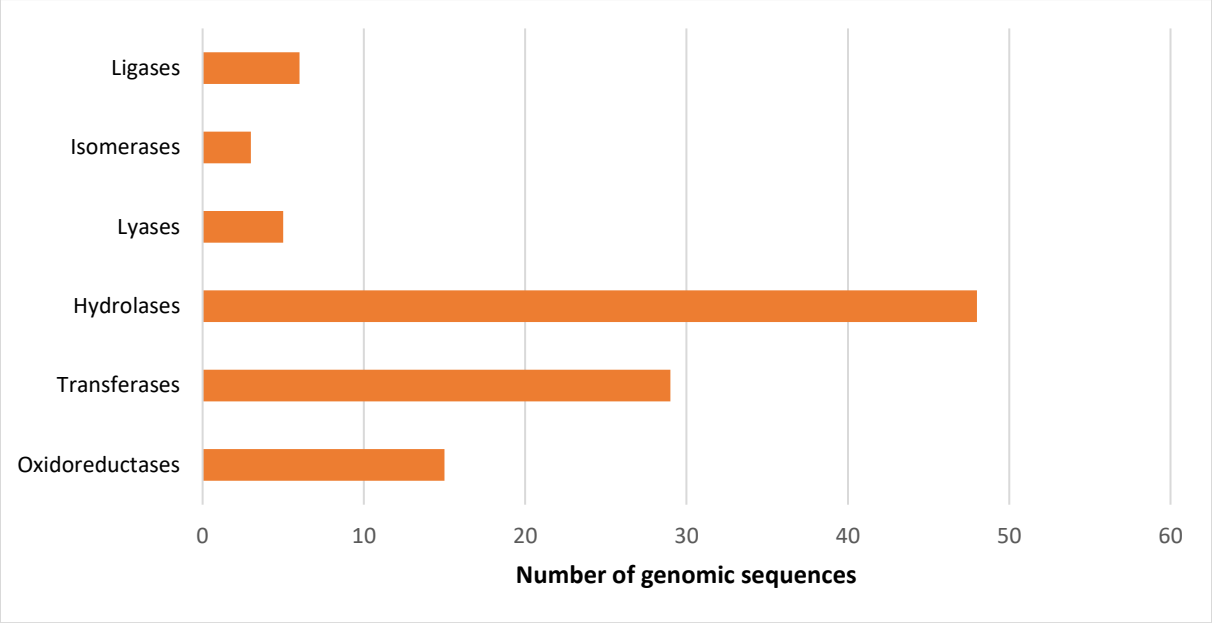

Figure 24. KEGG ontology distribution of the major enzymes involved in metabolism of secondary compounds in *C. sinensis*.

Among the 65 pathways, caffeine (Figure 25) and phenylalanine pathways (Figure 26) that are directly involved in caffeine, theobromine and flavonoid biosynthesis were identified.

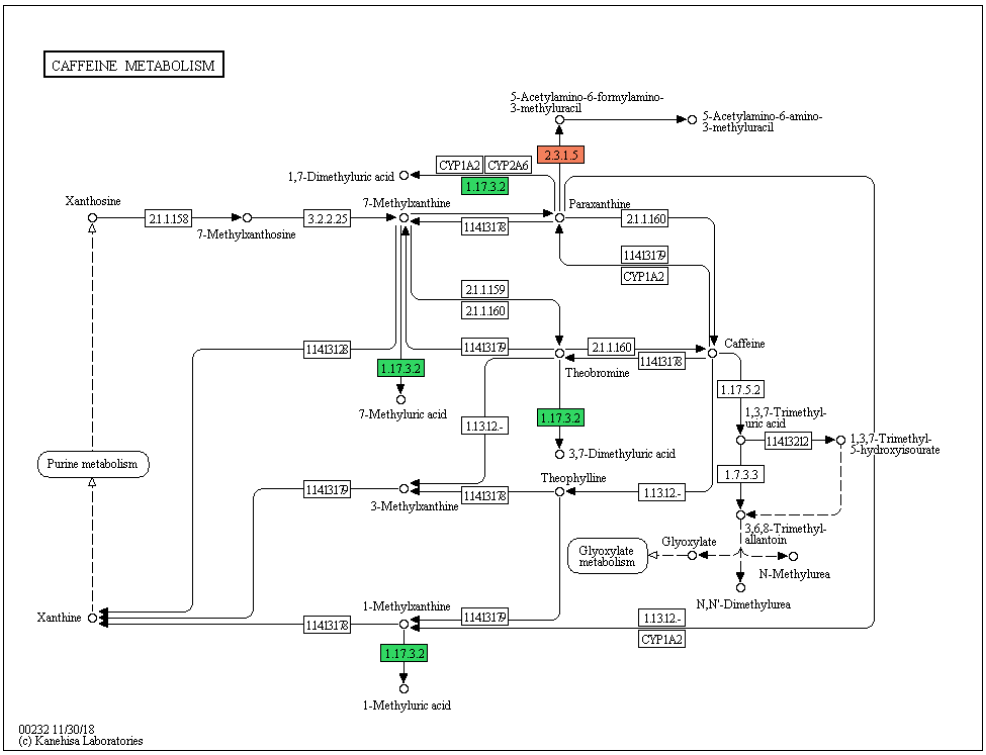

Figure 25. Caffeine and theobromine biosynthetic pathway found in the 2779 SNP sequence tags are described by the different coloured ECs (one color for each EC).

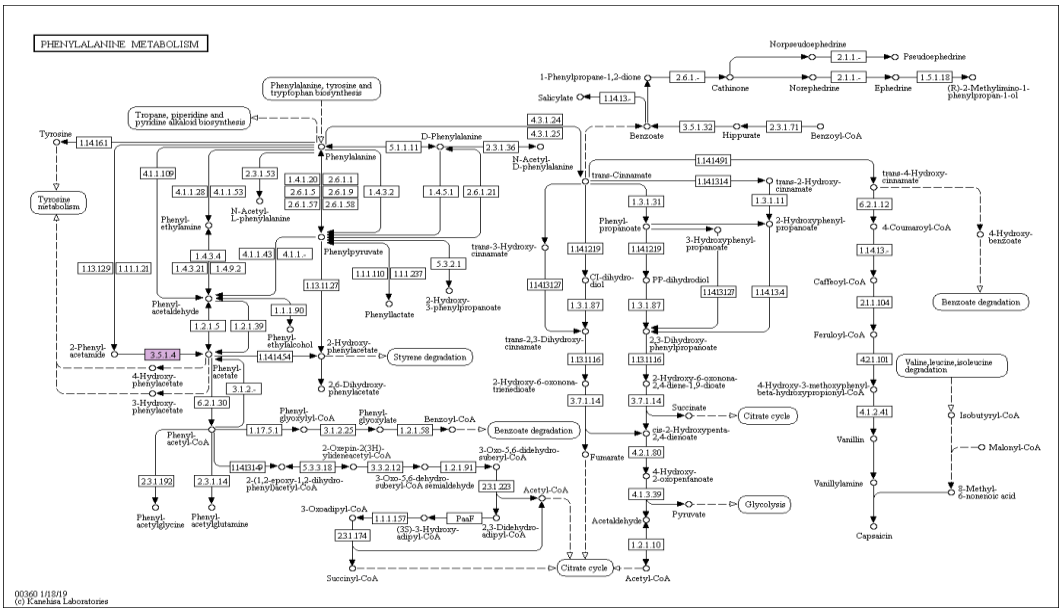

Figure 26. Phenylalanine metabolism pathway identified in the 2779 SNP sequence tags.

7.3.7. Markers significantly associated with the biochemical compounds

The BLAST results for all the markers that were significantly associated with the biochemical traits at R^2>0.2, their gene ID, p-values, scaffold position, the proportion explaining variation by the marker (R^2) and predicted gene function are presented on Table 18. In total, 64 SNP markers were significantly associated with the biochemical traits (Table S7. 4.). After further analysis, 44 SNP markers with R^2>0.2 were selected and the candidate genes associated with the markers identified in the tea plant information archive database (http://tpia.teaplant.org/FunctionalAnnotation_Locus.html) (Xia et al., 2019) (Table 18).

Twenty-three SNP markers were significantly (p < 3.59E-04) associated with ECG. TP71141 on scaffold Scaffold3639 was the most significant marker for ECG explaining 42% of the variation (Table 18). Twelve markers were significantly associated with GCG, and TP63459 (Scaffold4860) explained the most variation at 48% (Table 18). Five markers were significantly (p< 3.59E-04) associated with theogallin. TP39815 on scaffold Scaffold913 was the most significant marker for theogallin and it explained 37% of the trait variation. The SNP marker TP394 (Scaffold261) was significantly (p< 3.59E-04) associated with caffeine and theobromine. Additionally, markers TP71786 (Scaffold2336) and TP16595 (Scaffold629) were significantly (p < 3.59E-04) associated with EGC and GA, respectively (Table 18).

Table 18. SNP marker loci that are significantly (p < 3.59E-04) associated with the traits and the proportion of phenotypic variation they explain (R^2).

Trait	Marker	p	R2	Gene ID	Gene	Scaffold	Nearest predicted gene function
Caffeine	TP394	8.99E-05	0.20	TEA028063.1	LOC100250742	Scaffold261	NADH dehydrogenase (ubiquinone) Fe-S protein 4
ECG	TP39815	2.29E-05	0.34	TEA032995.1	LOC105796523	Scaffold913	LRR receptor-like serine/threonine-protein kinase GSO2 isoform X
	TP59981	2.24E-05	0.23	TEA033170.1	LOC100245826	Scaffold671	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase isoform X1
	TP71141	6.57E-06	0.42	TEA012194.1	LOC100264944	Scaffold3639	exopolygalacturonase
	TP77023	0.000286	0.20	TEA019659.1	LOC104594823	Scaffold3740	fasciclin-like arabinogalactan protein 11
	TP72124	1.71E-05	0.26	TEA030042.1	LOC18592248	Scaffold319	Histone acetyltransferase of the MYST family 1 isoform 2
	TP45007	1.13E-05	0.37	TEA014458.1	LOC112526886	Scaffold41	hypothetical protein Ccrd_023771
	TP36831	1.32E-06	0.30	TEA032612.1	LOC101214471	Scaffold3498	hypothetical protein Csa_5G146940
	TP75665	2.24E-05	0.23	TEA026342.1	LOC18788208	Scaffold1492	hypothetical protein PRUPE_ppa022614mg
	TP49950	2.12E-07	0.29	TEA002229.1	ycf1	Scaffold1983	hypothetical protein VITISV_018091
	TP81945	0.000127	0.22	TEA008015.1	ycf1	Scaffold275	hypothetical protein VITISV_029931
	TP62088	0.000102	0.21	TEA001150.1	LOC103837981	Scaffold671	leucine-rich repeat extensin-like protein 4
	TP47049	3.4E-05	0.28	TEA010521.1	LOC110008586	Scaffold838	mechanosensitive ion channel protein 5-like
	TP28250	2.21E-05	0.23	TEA016188.1	LOC102596302	Scaffold411	mitochondrial phosphate carrier protein 1
	TP19784	2.71E-06	0.23	TEA016025.1	LOC104606768	Scaffold104	nifU-like protein 4
	TP75737	0.000214	0.21	TEA029957.1	LOC100250223	Scaffold251	nuclear pore complex protein NUP160
	TP59980	2.24E-05	0.23	TEA022841.1	LOC111925338	Scaffold981	poly(rC)-binding protein 3-like isoform X1
	TP72125	9.98E-06	0.29	TEA018384.1	LOC100257644	Scaffold943	putative ATP-dependent helicase HRQ1 isoform X3
	TP12308	0.000156	0.23	TEA015664.1	LOC105782641	Scaffold3726	transmembrane 9 superfamily member 1-like isoform X1
	TP14758	8.97E-05	0.24	TEA021748.1		Scaffold214	uncharacterized protein
	TP8361	0.000282	0.32	TEA021225.1	LOC102594196	Scaffold4319	uncharacterized protein LOC102594196
	TP13032	1.26E-05	0.37	TEA029936.1	LOC102598318	Scaffold251	unnamed protein
	TP67837	6.26E-07	0.34	TEA018961.1		Scaffold242	unnamed protein product
	TP81402	9.25E-05	0.24	TEA002759.1		Scaffold2800	unnamed protein product
EGC	TP71786	4.45E-05	0.20	TEA022074.1	LOC105173585	Scaffold2336	uncharacterized protein LOC105173585 isoform X2
GA	TP16595	8.43E-05	0.20	TEA008540.1	LOC104437397	Scaffold629	probable mitochondrial chaperone BCS1-B
GCG	TP53892	5.2E-05	0.24	TEA009955.1	LOC100251534	Scaffold5952	ATP-dependent zinc metalloprotease FtsH
	TP63459	3.36E-05	0.48	TEA007457.1	LOC114293921	Scaffold4860	cytochrome P450 704C1-like protein
	TP957	0.000151	0.22	TEA003188.1	LOC100241573	Scaffold930	E3 ubiquitin protein ligase RIN2
	TP71141	1.78E-04	0.30	TEA012194.1	LOC100264944	Scaffold3639	exopolygalacturonase
	TP15601	0.000233	0.21	TEA020725.1	LOC100266022	Scaffold1672	glutamic acid-rich protein
	TP45007	2.11E-04	0.35	TEA014458.1	LOC112526886	Scaffold41	hypothetical protein Ccrd_023771
	TP47049	1.32E-04	0.23	TEA010521.1	LOC110008586	Scaffold838	mechanosensitive ion channel protein 5-like
	TP28250	7.97E-05	0.20	TEA016188.1	LOC102596302	Scaffold411	mitochondrial phosphate carrier protein 1
	TP16595	8.65E-06	0.26	TEA008540.1	LOC104437397	Scaffold628	probable mitochondrial chaperone BCS1-B
	TP27072	7.53E-05	0.36	TEA003563.1	LOC104241249	Scaffold2924	probable receptor protein kinase TMK1
	TP69675	0.000142	0.22	TEA011118.1	LOC100527400	Scaffold879	uncharacterized protein LOC100527400
	TP67837	1.18E-04	0.21	TEA018961.1		Scaffold242	unnamed protein product
Theobromine	TP394	7.61E-06	0.26	TEA028063.1	LOC100250742	Scaffold261	NADH dehydrogenase (ubiquinone) Fe-S protein 4
Theogallin	TP39815	2.49E-05	0.37	TEA032995.1	LOC105796523	Scaffold913	LRR receptor-like serine/threonine-protein kinase GSO2 isoform X
	TP71141	7.33E-05	0.28	TEA012194.1	LOC100264944	Scaffold3639	exopolygalacturonase
	TP19784	1.10E-05	0.20	TEA016025.1	LOC104606768	Scaffold104	nifU-like protein 4
	TP72125	3.06E-04	0.21	TEA018384.1	LOC100257644	Scaffold943	putative ATP-dependent helicase HRQ1 isoform X3
	TP67837	6.50E-06	0.29	TEA018961.1		Scaffold242	unnamed protein product

7.3.8. Heritability of the biochemical traits

The heritabilities of all the biochemical traits were calculated from both phenotypic and genomic data. Broad-sense heritabilities of the traits were estimated using the procedure outlined by Henderson (1975), while all markers were incorporated to estimate the genomic heritability (GBLUP) using a realized genomic relationship matrix as described by VanRaden (2008). Broad sense heritability ranged from ECG (0.67) to EC (0.09). Traits with high broad sense heritability were ECG (0.67) and EGC (0.65) (Table 19). Traits with low broad sense heritability were EC (0.09) catechin (0.21), theanine (0.24), GA (0.24) and theobromine (0.28) (Table 19). Genomic heritability ranged from 0.99 (Theogallin) to 0.52 (EC) (Table 19). Traits with high genomic heritabilities were theogallin (0.99), ECG (0.99), theobromine (0.95), EGC (0.92) and EGCG (0.92) (Table 19). Traits with low genomic heritability were EC (0.52) and theanine (0.59) (Table 19). The correlation between broad sense and genomic heritability was 0.67.

Table 19. Broad sense and genomic heritability estimates.

Trait	Broad sense heritability	Genomic heritability
Caffeine	0.43	0.8419448
catechin	0.21	0.7006664
EC	0.09	0.5287559
ECG	0.67	0.9999315
EGC	0.65	0.9255677
EGCG	0.43	0.9246394
GA	0.24	0.7165658
GC	0.59	0.7958524
GCG	0.39	0.7416647
Theanine	0.24	0.5924403
theobromine	0.28	0.9551852
Theogallin	0.34	0.9999318

All traits had a higher genomic heritability compared to broad sense heritability (Figure 27).

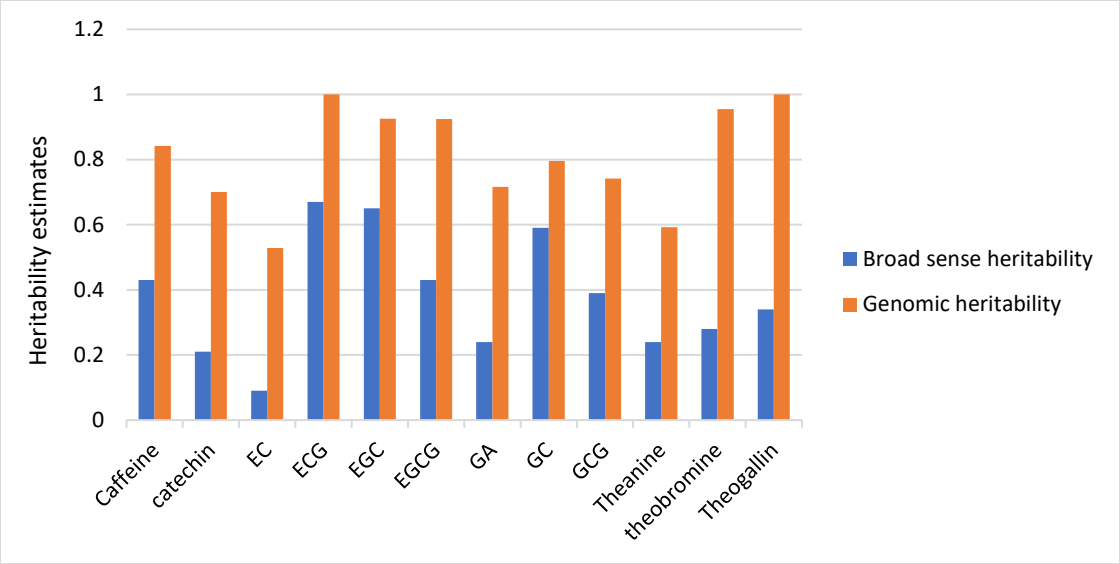

Figure 27. Comparison of broad sense and genomic heritability.

7.3.9. Prediction accuracies for the biochemical traits

The 2779 SNPs were regressed to each of the 12 biochemical traits and GEBVs were estimated for all the genotypes using five different GS prediction models: RRBLUP (Endelman, 2011; Meuwissen et al., 2001), BayesLASSO (Usai et al., 2009), BayesA (Meuwissen et al., 2001) BayesB (Meuwissen et al., 2001) and BayesCπ (Meuwissen et al., 2001). Correlations of the prediction accuracies were calculated within the 103-training population using 5-fold cross-validation approach and repeated 10 times.

For theobromine, the models that gave the highest accuracies were BayesLASSO (0.65) and RRBLUP (0.65). The least accuracy was obtained using BayesB (0.52) (Table 20). RRBLUP (0.58) had the highest prediction accuracy for caffeine, while BayesB (0.43) had the lowest prediction accuracy for the same trait (Table 20). TP394 was the most significant marker for caffeine and theobromine and it explained 20% and 26% of the total variation, respectively (Table S7.4). For theogallin, RRBLUP (0.61) and BayesA (0.61) models performed the best, while BayesCπ (0.57) had the lowest prediction accuracy. The most significant marker for theogallin was TP39815 and it explained 37% variation (Table S7.4). RRBLUP (0.44) had the highest prediction accuracy for GA, while BayesB (0.37) had the lowest. RRBLUP (0.62) and BayesA (0.62) had the highest prediction accuracies for both EC and ECG. Similarly, for both

ECG and EC, BayesLASSO (0.61), BayesCπ (0.61) and BayesB (0.61) had the lowest prediction accuracy (Table 20). The marker TP71141 was the most significant for ECG and it explained 42% of the variation (Table S7.4). For EGC, BayesLASSO (0.56) had the highest prediction accuracy, while BayesCπ (0.54) had the lowest. For GC, BayesB (0.36) had the highest prediction accuracy while RRBLUP (0.31) and BayesA (0.31) had the lowest. (Table 20). For GCG, BayesLASSO (0.50), had the highest prediction accuracy, while BayesB (0.44) recorded the lowest accuracy (Table 20). The significant marker for GCG that explained the most variation was TP63459 at 48% (Table S7.4). The models with the highest prediction accuracy for EGCG were BayesB (0.64) and BayesCπ (0.64), while the lowest prediction accuracy was recorded by RRBLUP (0.61) and BayesA (0.61) (Table 20). For catechin, BayesB (0.4) had the highest prediction accuracy, while RRBLUP (0.33) had the lowest (Table 20). Two markers; TP32223 and TP32224 were the most significant for catechin and they each explained 17% of the total variation (Table S7.4). For theanine, BayesB, BayesBCπ and BayesLASSO had the highest prediction accuracy of 0.63 each. However, RRBLUP (0.61) had the lowest prediction accuracy for theanine (Table 20).

The mean accuracies of the traits were averaged for all the GS models and the traits with the highest prediction accuracy were; ECGG (0.62), Theanine (0.62), EC (0.61), ECG (0.61) and theobromine (0.61) (Table 20). Traits with the lowest GS prediction for all the models were GC (0.33), catechin (0.36) and GA (0.42) (Table 20). Similarly, the mean GS accuracies for all the traits was calculated across all the traits. The performance of all the GS models were almost the same, with RRBLUP (0.53), BayesLASSO (0.53) and BayesA (0.53) performing slightly better than the other models (Table 20). BayesB had the lowest prediction accuracy in majority of the traits.

Table 20. Mean accuracy of traits for the five studied GS models.

GS model	TB	Caffeine	Theogallin	GA	EC	GC	GCG	ECG	EGC	EGCG	Catechin	Theanine	Mean
BayesA	0.64	0.56	0.61	0.43	0.62	0.31	0.49	0.62	0.55	0.61	0.35	0.62	0.53
BayesB	0.52	0.43	0.59	0.37	0.61	0.36	0.44	0.61	0.55	0.64	0.40	0.63	0.51
BayesCπ	0.60	0.48	0.57	0.40	0.61	0.33	0.48	0.61	0.54	0.64	0.37	0.63	0.52
BayesLASSO	0.65	0.54	0.58	0.43	0.61	0.32	0.50	0.61	0.56	0.62	0.37	0.63	0.53
RRBLUP	0.65	0.58	0.61	0.44	0.62	0.31	0.49	0.62	0.55	0.61	0.33	0.61	0.53
Mean	0.61	0.52	0.59	0.42	0.61	0.33	0.48	0.61	0.55	0.62	0.36	0.62	

7.4. Discussion

7.4.1. Population structure

The genotypes used in this study were selected from a larger population that was crossed and planted at two sites in UTK. The training population was selected to be as unrelated as possible from all the available genotypes in UTK breeding programme to create a diverse training population for use in genomic prediction. The genetic diversity in the 103 genotypes was consequently large, which is important when constructing GS prediction models. In this study, no major population structure was detected, signifying that genetic diversity between alleles was large. The absence of strong genetic relationships across the entire population also confirms the highly heterogenous and heterozygous nature of tea (*Camellia sinensis* (L.) O. Kuntze) (Wei et al., 2018b; Wachira et al., 2005). A few studies on genetic relationships have been conducted in tea using SSR markers (Wambulwa et al., 2016) and SNPs (Meegahakumbura et al., 2018a). They concluded that the main varieties of tea grown in East Africa include *Camellia sinensis* var. *sinensis* and *Camellia sinensis* var. *assamica*. *Camellia sinensis* var. *assamica* is the main type under cultivation and has contributed more in major tea breeding programmes in East Africa (Wambulwa et al., 2016). This study utilised SNP markers generated using GBS because they are the most abundant and contribute to phenotypic variation for many traits (Yang et al., 2018a). The SNP markers used in this study revealed great genetic diversity among the genotypes.

7.4.2. Repeatability

Repeatability is also defined as broad-sense heritability and is the proportion of phenotypic variance explained by heritable (additive) and nonheritable (dominance, epistasis) genetic variance (Kruijer et al., 2015). The present results show that repeatability was high for all the traits except catechin and theobromine, indicating that the NMR spectroscopy used was reliable in determining the absolute values of the biochemical compounds. The low estimates of repeatability for catechin indicated low genetic variance and suggest differences in its response to the environment. This indicates that more replicates are

required to give a reliable estimate and selection on the basis of first record is not recommended for catechin. Yuan et al. (2014) while quantifying caffeine, gallic acid, theanine, epicatechin, EGC, ECG and EGCG using ^{1}H-NMR spectroscopy reported high accuracy, precision and repeatability values. They recommended ^{1}H-NMR spectroscopy to be adopted as a simple, rapid, and powerful tool for tea quality evaluated (Yuan et al., 2014).

7.4.3. Markers and candidate genes associated with tea quality

Tea quality is influenced by catechins, caffeine, amino acids and aroma compounds present in fresh tea shoots (Owuor et al., 2007; Balentine et al., 1997). These compounds influence the taste, mouth feel, colour and aroma of tea (Scharbert et al., 2004; Liang et al., 2003). However, tea quality is a complex trait that is influenced by variety, processing conditions, post-harvest handling and environmental conditions (Borse, 2012). This study focused on identification of significant SNPs and functional annotation of the candidate genes associated with tea quality. Similarly, we assigned biosynthetic pathways associated with biochemical compounds influencing tea quality.

The presence of high levels of repeat regions and scaffolds still present in both draft tea genomes is a major problem for studies involving genome wide association studies (GWAS) in tea (Wei et al., 2018b; Xia et al., 2017c) . In this study, only 929 SNP markers (33.42% of the total GBS tag sequences) were mapped to the draft tea genome, implying that the genome assembly and content of the current tea genome is incomplete; implying that further sequencing is necessary to improve the genome assembly and content of the current tea genome. Only 217 sequences were associated with an annotation, suggesting that while the PstI restriction endonuclease used may be enriching for coding sequence and low copy sequences, it is not exclusive to gene regions or that existing annotation is limited. Given that in silico detection is based on only 64bp, it is perhaps not surprising as a result. However, it does imply that the annotations and sequence content of the two draft genomes is incomplete.

GWAS is a useful tool for identifying the genetic loci and identifying candidate genes responsible for genetic variation underlying targeted quantitative traits (Korte et al., 2013b). The accuracy of GWAS

to identify markers associated with the traits of interest relies on the marker density, population size, genetic architecture of the trait, and statistical technique used (Zhang et al., 2015). In this study, the MLM model using the biochemical data and 2779 SNP markers were used to search for significant SNP markers associated with each biochemical trait. The MLM model was used because it is more stringent compared to the GLM model and allows for a reduction in spurious associations by simultaneously estimating and incorporating the population structure and the kinship among individuals in the analysis (Su et al., 2019). Population sizes ranging from 50 to 300 individuals have commonly been used in GWAS (Su et al., 2019; Zhou et al., 2017). However, several factors such as cost of phenotyping, plant species size and breeding objectives influence the population size.

The main genes involved in flavonoid and alkaloid biosynthesis in plants have been characterized (Yue et al., 2018). Flavonoids are important polyphenols in tea and are grouped into catechins, flavones, flavonols, flavanones, anthocyanins and isoflavonoids (Hodgson et al., 2010). Flavonoids are synthesized through the phenylpropanoid biosynthetic pathways, whereby phenylalanine is first converted into 4-coumaroyl-CoA, which goes into the flavonoid biosynthesis pathway (Wei et al., 2011; Guo et al., 2017). The phenylalanine pathway was identified in this study. The main step in the flavonoid biosynthetic pathway involves chalcone formation, and the chalcone synthase gene (CHS) encodes the key enzyme involved (Dare et al., 2013). In this study, SNP marker TP2350 was aligned to one of the genes involved in phenylalanine metabolism namely acylamidase (EC:3.5.1.4). Additionally, several groups of enzymes were identified; isomerases (ec:5.3.1.5) by SNP TP22681, reductases (ec:1.1.1.169) by SNP TP59880 and hydroxylases (ec:3.6.1.15, ec:3.4.16.4) by 25 different SNP markers. These enzymes modify the basic flavonoid skeleton (Martens et al., 2010), leading to the different categories of catechins.

Transferases modify the flavonoid backbone with sugars, methyl groups and acyl moieties (Ciarkowska et al., 2019; Falcone Ferreyra et al., 2012) hence modulating their physiological activity by altering the solubility, reactivity and interaction of flavonoids with cellular targets (Bowles et al., 2005). Acyltransferases are involved in the biosynthesis of anthocyanidins and various flavonoid groups (Chen et al., 2011b). Glycerol-3-phosphate acyltransferase gene was identified by SNP marker TP12852. The gene

UDP: flavonoid glycosyltransferase was associated with SNP markers TP52078 and TP82001. Flavonoids are glycosylated with pentoses and hexoses, resulting in the functional and structural diversity of flavonoids (Jones et al., 2001). Flavonoid glycosylation reactions mainly depend on UDP-sugar dependent glycosyltransferases that use UDP sugars as sugar donors (Lairson et al., 2008). Diacylglycerol O-acyltransferase (EC:2.3.1.20) and amino-acid N-acetyltransferase (EC:2.3.1.1) were also identified by SNP markers TP21005 and TP68175, respectively in this study. Similarly, the KEGG pathway analysis showed hypoxanthine oxidase (EC:1.17.3.2) and N-acetyltransferase (EC:2.3.1.5) as the main enzymes identified in caffeine and theobromine biosynthesis, and were identified by SNP markers TP45290, TP4662 and TP68175. Jin et al. (2016) reported tea caffeine synthase (TCS) as the main enzyme involved in the formation of caffeine in tea by methylation of N-3(theobromine synthase, TS, EC 2.1.1.159) and N-1 (caffeine synthase, CS, EC 2.1.1.160).

The only significant marker for caffeine and theobromine was TP394 on scaffold Scaffold261 and it was nearest to the gene that encodes NADH dehydrogenase (ubiquinone) Fe-S protein 4. NADH dehydrogenase (ubiquinone) Fe-S protein 4 enzymes recycle NAD+ for use in glycolysis and the TCA cycle and transfer the resulting electrons onto ubiquinone in the mitochondrial inner membrane (Galkin et al., 2006). Among the alkaloids, caffeine is the most abundant in tea, while theophylline and theobromine are present in low concentrations (Ahmad Bhawani et al., 2015; Chen et al., 2010). Caffeine and theobromine are important quality indicators in tea, as they influence the taste and colour of tea (El-Shahawi et al., 2012b).

SNP marker TP63459 on scaffold Scaffold4860 explained the most variation for the GCG trait and was near the gene encoding cytochrome P450 704C1-like protein. Cytochrome P450 enzymes catalyze hydroxylation reactions and play critical roles in the biosynthesis of secondary metabolites such as flavonoids, terpenoids, alkaloids, fatty acids, hormones and pigments (Schuler, 1996). In tea (*Camellia sinensis* (L.) O. Kuntze), Eminoğlu et al. (2018) characterized a putative cytochrome P450 monooxygenase gene (Csp450). However, this gene did not show preharvest period expression for the studied months; May, July and September. They recommended further investigations and functional characterization of this gene

as it could reveal the function of the gene in the tea plant metabolism (Eminoğlu et al., 2018). E3 ubiquitin protein ligase has been profiled in tea *(Camellia sinensis* (L.) O. Kuntze) and is involved in catechins biosynthesis (Xie et al., 2019). They reported that E3 ubiquitin protein ligases are enriched in the chalcone isomerase activity and flavonoid metabolic processes. Analysis of KEGG pathway also revealed that E3 ubiquitin proteins ligases were mapped to similar pathways including flavonoid biosynthesis, phenylpropanoid biosynthesis, ribosome and protein processing in endoplasmic reticulum (Xie et al., 2019). ATP-dependent zinc metalloprotease FtsH on Scaffold5952 was associated with GCG and was identified by SNP TP53892. ATP-dependent zinc metalloprotease FtsH enzyme is known to be expressed during the withering process of tea *(Camellia sinensis (L.) O. Kuntze*) (Wu et al., 2018). Withering is the loss of moisture and is a critical step in the manufacture of black tea as it significantly impacts on its' quality (Deb et al., 2016).

ECG and theogallin were associated with SNP marker TP71141 on Scaffold3639, and exopolygalacturonase enzymes, which are a group of pectinases that have been used for fermentation of tea and to improve tea quality *(Camellia sinensis* (L.) O. Kuntze) (Thakur et al., 2012). Fermentation is an important step in tea manufacture and is catalyzed by the enzyme polyphenol oxidase (Samanta et al., 2015). Bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase was associated with ECG on scaffold Scaffold671 and has been found to have a prenate transaminase activity, which is a direct precursor for the biosynthesis of the aromatic amino acids phenylalanine and tyrosine (de la Torre et al., 2014). Phenylalanine is involved in the biosynthesis of phenylpropanoid, which is a major amino acid in the biosynthesis of catechins (Wei et al., 2011).

7.4.4. Genomic prediction models and heritability

GS simultaneously estimates all loci, haplotype and marker effects across the entire genome to produce an estimate of the GEBVs of individuals and it requires that every marker is in linkage disequilibrium with every QTL affecting the trait of interest (Heffner et al., 2009a). In this study, five GS models namely; RRBLUP, BayesA, BayesB, BayesCπ and BayesLASSO were fitted to predict the twelve

biochemical traits. GS models attempt to captures total additive genetic variance across the entire genome to estimate GEBVs among the selection candidates based on the sum of all marker effects (Lorenz et al., 2011a). RRBLUP assumes marker effects are equally shrunk and all markers have equal variance (Meuwissen et al., 2001). The Bayesian models assume the markers have different amounts of variation and are more flexible while predicting traits with different genetic architectures (Habier et al., 2011). Bayesian models are therefore suited for traits that are controlled by few large-effect genes compared to RRBLUP (Beaulieu et al., 2014; Meuwissen et al., 2001).

Biochemical traits are considered to be quantitative traits and are controlled by many genes (Kamunya et al., 2010). Therefore, differences between the prediction accuracy of the different models was not expected to be large. In this study, RRBLUP, BayesLASSO and BayesA slightly outperformed BayesB and BayesCπ. In oil palm (Kwong et al., 2017) and cassava (Wolfe et al., 2017), the range in GS accuracies was between 0.31-0.32 and 0.30-0.33 respectively, and machine learning methods slightly outperformed the other methods. In potatoes, Sverrisdóttir et al. (2017) found that all the models performed equally at 0.56. However, most GS studies in forest trees reveal that RRBLUP outperformed all the other models, implying that all major traits in forest trees fit the infinitesimal model (Grattapaglia et al., 2018; El-Dien et al., 2015; Resende et al., 2012b). Wang et al. (2015) reported BayesCπ outperformed other models where a small number of loci had a large effect on a trait, while there were no differences among the models when the trait was controlled by many minor genes in wheat. However, RRBLUP model (Endelman, 2011; VanRaden, 2008; Meuwissen et al., 2001) could be implemented in tea because it is computationally simple to fit compared to the other models and has outperformed other models in many crops (Jan et al., 2016).

The heritability of each trait was estimated from the replicated phenotypic data and from genetic data (de los Campos et al., 2015). The heritability of a trait significantly affects the response to selection and improves the efficiency of GS over phenotypic selection (Zhang et al., 2017b; Hayes et al., 2009b). High heritability leads to increased gain from selection for the traits of interest (Kruijer et al., 2015; de los Campos et al., 2015). Broad sense and genomic heritability estimates were different for each trait, ranging from high to low values. For ECG and EGC, the heritability was high for both methods, suggesting that

little genetic gains could be achieved for this trait using SNP markers. Genomic heritability estimates were higher than those from the phenotypic heritability for all the traits, suggesting that higher genetic gains can be achieved using SNP markers. This illustrates that using molecular markers in breeding programmes could improve the accuracy of obtaining reliable breeding values. The genomic heritability is the proportion of phenotypic variance explained by the regressing phenotypes on molecular markers. Many polymorphic markers are required to accurately estimate relatedness especially for distant relatives. RRBLUP relies on estimating the realized kinship and is more accurate in estimating the hereditary relationships among genotypes (de Roos et al., 2009). Our results agreed with other studies that GS is more beneficial where traits have a low heritability (Goddard, 2009). For instance, EC, catechin, theanine, theobromine and GA had low heritabilities but had good GS accuracies. This further shows that additional genetic gains could be achieved using GS for traits with low heritabilities. For traits with low heritabilities, many markers and a large training population is required to obtain higher GS accuracies as suggested by Lorenz et al. (2011b). Heritability could also be improved by increasing the number of replications, years of recording phenotypic data and experimental sites (Zhang et al., 2017b). Therefore, when designing a GS pipeline heritability of the target traits in the training population should be high to achieve a high GS accuracy.

7.4.5. Implementing GS in tea breeding

Generally, our results suggest that GS has a great potential in predicting the performance of tea. The predictions were higher for specific traits namely ECGG, theanine, EC, ECG and theobromine. The main challenge facing all tea breeding programmes is the long generation interval, as it takes between 3 to 6 years for tea to grow from seedling to flowering (Mondal, 2014a). This means that developing an improved tea variety using conventional methods requires many years of field selection (Corley et al., 2018). The greatest impact of GS in tea breeding is reducing the length required to develop improved varieties by bypassing some of the field-testing steps. This could be done by first applying GS early at the nursery stage. The genotypes with high breeding values could then be tested in the field, selected and released for commercial planting. Compared to conventional field selection method, GS can improve

genetic gain per unit time significantly. The main factors that could be considered before implementing GS in a tea breeding programme include prediction models, the size of the training population, the relationship between the training and the breeding populations, heritability, genetic architecture of the trait of interest in tea, marker density and cost-effective genotyping platforms.

The training population used to construct GS model should be closely related to the breeding population and should be large as possible as this improves the accuracy of estimating marker effects (Lorenz et al., 2011b). Zhang et al. (2017b) showed that prediction accuracy increased for all the traits in maize with increasing training population size. Since tea has a high allelic diversity, the training population should consist of genotypes with broad genetic diversity for the traits of interest.

Trait heritability is a key factor that significantly impacts on the accuracy of genomic selection (Heffner et al., 2011). Our findings agreed with previous studies that prediction accuracy increases with an increase in trait heritability (Zhang et al., 2017b). However, heritability could be improved by designing field experiments for the training population to increase the number of replications, testing sites and years of data collection (Mackay et al., 1999).

The density and type of markers to be used in constructing GS models influence the prediction accuracy (Goddard et al., 2011). In this study, SNP markers were used because they are abundant in the plant genome and they give higher prediction accuracies compared to other markers (Kwong et al., 2017). Cheaper options of SNP genotyping include GBS, a simple highly-multiplexed next generation sequencing platform that generates large numbers of SNPs (Elshire et al., 2011). GBS is less expensive compared to other platforms and can provide genome-wide marker coverage for species that lack a reference genome (Davey et al., 2011). However, SNP markers obtained by GBS usually contain a large proportion of missing data across samples because fragments of the genome are sequenced at low depth, and hence some loci could have zero coverage (Elshire et al., 2011). In GS, using a large number of markers and selecting a suitable imputation algorithm enables the use of low-density SNP markers without a major loss in prediction accuracy (Mulder et al., 2012; Habier et al., 2009). The most common imputation algorithms that could be used include; mean, singular value decomposition (SVD), traditional k nearest neighbor (kNN), expectation

maximization (EM) and random forest regression imputation algorithms (Rutkoski et al., 2013; Marchini et al., 2010). GS requires genome wide markers that explain most genetic variation (Meuwissen et al., 2001) and every marker should be in linkage disequilibrium with every gene affecting the trait of interest (Hayes et al., 2009b). Therefore an increase in the length of LD or in marker number steadily improves the prediction accuracy (Asoro et al., 2011a).

The type of model used for GS could impact on the prediction accuracy and mainly depend on the complexity of the trait (Crossa et al., 2017). The main GS models developed differ on assumptions of the trait architecture and they include RRBLUP, GBLUP, Reproducing Kernel Hilbert Spaces(RKHS), Bayesian models (BayesA, BayesB, BayesC, BayesLASSO) and machine learning (Wang et al., 2018b; Lorenz et al., 2011a). A suitable model could be tested and selected based on the complexity of the trait.

7.5. Conclusion

In this study, the 2779 SNPs were used in a GWAS to identify significant markers and candidate genes associated with quality traits in tea. We also used a variety of models to predict tea quality using catechins and caffeine present in fresh green leaf. Our study proves that conducting GWAS using SNP markers represent a powerful approach for dissecting tea quality and identifying candidate genes in tea. We identified 64 significant markers and 44 candidate genes associated with caffeine, ECG, GCG, theogallin, GA and theobromine.

The main objective of GS is to use large numbers of molecular markers to estimate breeding values of a population via model construction and then subsequent selection by genotypes, rather than phenotypes in the next generation. Our results showed that the differences in accuracies between the methods evaluated were small. Generally, RRBLUP, BayesLASSO and BayesA models slightly outperformed the other methods. However, RRBLUP could be selected for use because it is computationally simple to use. The high GS accuracies for nearly all the traits from our results clearly demonstrates the potential of GS using genome wide SNP markers to predict high quality varieties in a tea breeding programme. While the main benefit of GS in tea breeding is expected to be the reduction of the breeding cycle length by several years,

the use of a realized genomic relationship matrix also enables the precise evaluations of genetic relationship and heritabilities.

References

Abdel-Rahman A, Anyangwe N, Carlacci L, Casper S, Danam RP, Enongene E, et al. (2011). The safety and regulation of natural products used as foods and food ingredients. *Toxicological Sciences, 123*(2), 333-348.

Abdurakhmonov IY, Abdukarimov A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International journal of plant genomics, 2008*.

Ahmad Bhawani S, Fong SS, Mohamad Ibrahim MN. (2015). Spectrophotometric Analysis of Caffeine. *International Journal of Analytical Chemistry, 2015*, 1-7. doi:10.1155/2015/170239

Ahmed S, Orians CM, Griffin TS, Buckley S, Unachukwu U, Stratton AE, et al. (2014). Effects of water availability and pest pressures on tea (Camellia sinensis) growth and functional quality. *AoB Plants, 6*.

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, et al. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and applied genetics, 113*(8), 1409-1420.

Akuli A, Pal A, Bej G, Dey T, Ghosh A, Tudu B, et al. (2016). A MACHINE VISION SYSTEM FOR ESTIMATION OF THEAFLAVINS AND THEARUBIGINS IN ORTHODOX BLACK TEA. *International Journal on Smart Sensing & Intelligent Systems, 9*(2).

Alasalvar C, Topal B, Serpen A, Bahar B, Pelvan E, Gökmen V. (2012). Flavor characteristics of seven grades of black tea produced in Turkey. *Journal of Agricultural and Food Chemistry, 60*(25), 6323-6332.

Alfatni MSM, Shariff ARM, Shafri HZM, Saaed OMB, Eshanta OM. (2008). Oil palm fruit bunch grading system using red, green and blue digital number. *Journal of Applied Sciences, 8*(8), 1444-1452.

Anami BS, Burkpalli VC. (2009). Color based identification and classification of boiled food grain images. *International journal of food engineering, 5*(5).

Ananthacumaraswamy S, Singh ID. (2002). An overview on the factors affecting the quality of Sri Lankan black tea. *Journal of food science and technology, 39*(5), 449-457.

Arce L, Ríos A, Valcarcel M. (1998). Determination of anti-carcinogenic polyphenols present in green tea using capillary electrophoresis coupled to a flow injection system. *Journal of Chromatography A, 827*(1), 113-120.

Arfin T, Semangun P. (1999). Tea industry in Indonesia. *Aravali Book International Pvt Ltd, New Delhi*, 65–72.

Ariyarathna HACK, Gunasekare MTK, Kottawa-Arachchige JD, Paskarathevan R, Ranaweera KK, Ratnayake M, et al. (2011). Morpho-physiological and phenological attributes of reproductive biology of tea (Camellia sinensis (L.) O. Kuntze) in Sri Lanka. *Euphytica, 181*(2), 203-215. doi:10.1007/s10681-011-0399-9

Arshak K, Moore E, Lyons GM, Harris J, Clifford S. (2004). A review of gas sensors employed in electronic nose applications. *Sensor review, 24*(2), 181-198.

Ashihara H, Gillies FM, Crozier A. (1997). Metabolism of caffeine and related purine alkaloids in leaves of tea (Camellia sinensis L.). *Plant and cell physiology, 38*(4), 413-419.

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. (2011a). Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *The Plant Genome Journal, 4*(2). doi:10.3835/plantgenome2011.02.0007

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. (2011b). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The plant genome, 4*(2), 132-144.

Avise JC. (2012). *Molecular markers, natural history and evolution*: Springer Science & Business Media.

Awad H, Khamis MM, El-Aneed A. (2015). Mass spectrometry, review of the basics: ionization. *Applied Spectroscopy Reviews, 50*(2), 158-175.

Bae S-H, Suh H-J. (2007). Antioxidant activities of five different mulberry cultivars in Korea. *LWT-Food Science and Technology, 40*(6), 955-962.

Balasaravanan T, Pius PK, Kumar RR, Muraleedharan N, Shasany AK. (2003). Genetic diversity among south Indian tea germplasm (Camellia sinensis, C. assamica and C. assamica spp. lasiocalyx) using AFLP markers. *Plant Science, 165*(2), 365-372.

Balentine DA, Wiseman SA, Bouwens LCM. (1997). The chemistry of tea flavonoids. *Critical Reviews in Food Science & Nutrition, 37*(8), 693-704.

Banerjee B. (1992). Selection and breeding of tea. In *Tea* (pp. 53-86): Springer.

Baptista JAB, da P Tavares JF, Carvalho RCB. (1998). Comparison of catechins and aromas among different green teas using HPLC/SPME-GC. *Food Research International, 31*(10), 729-736.

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp .). *Plant Science, 242*, 23-36. doi:10.1016/j.plantsci.2015.08.021

Batley J, Edwards D. (2016). The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Current Opinion in Plant Biology, 30*, 78-81.

Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity, 113*(4), 343.

Bekele WA, Fiedler K, Shiringani A, Schnaubelt D, Windpassinger S, Uptmoor R, et al. (2014). Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions. *Plant, cell & environment, 37*(3), 707-723.

Ben-Ari G, Lavi U. (2012). Marker-assisted selection in plant breeding. In *Plant Biotechnology and Agriculture* (pp. 163-184): Elsevier.

Bensch S, Åkesson M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology, 14*(10), 2899-2914.

Bernardo R, Yu J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop science, 47*(3), 1082-1090.

Bhardwaj P, Sharma RK, Kumar R, Sharma H, Ahuja PS. (2014, 2014). *SSR marker based DNA fingerprinting and diversity assessment in superior tea germplasm cultivated in Western Himalaya*.

Bhattacharyya N, Bandyopadhyay R, Bhuyan M, Ghosh A, Mudi RK. (2005). Correlation of multi-sensor array data with 'tasters' panel evaluation for objective assessment of black tea flavour. *Int. Proc. ISOEN-2005, Barcelona, Spain*, 13-15.

Bhattacharyya N, Bandyopadhyay R, Bhuyan M, Tudu B, Ghosh D, Jana A. (2008). Electronic nose for black tea classification and correlation of measurements with "Tea Taster" marks. *IEEE transactions on instrumentation and measurement, 57*(7), 1313-1321.

Bhattacharyya N, Seth S, Tudu B, Tamuly P, Jana A, Ghosh D, et al. (2007). Monitoring of black tea fermentation process using electronic nose. *Journal of Food Engineering, 80*(4), 1146-1156. doi:10.1016/j.jfoodeng.2006.09.006

Bhuyan LP, Hussain A, Tamuly P, Gogoi RC, Bordoloi PK, Hazarika M. (2009). Chemical characterisation of CTC black tea of northeast India: correlation of quality parameters with tea tasters' evaluation. *Journal of the Science of Food and Agriculture, 89*(9), 1498-1507. doi:10.1002/jsfa.3614

Bian M, Skidmore AK, Schlerf M, Liu Y, Wang T. (2012). Estimating biochemical parameters of tea (Camellia sinensis (L.)) using hyperspectral technique. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39*, B8.

Bian M, Skidmore AK, Schlerf M, Wang T, Liu Y, Zeng R, et al. (2013). Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels. *ISPRS journal of photogrammetry and remote sensing, 78*, 148-156.

Blears MJ, De Grandis SA, Lee H, Trevors JT. (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology, 21*(3), 99-114.

Borchetia S, Das SC, Handique PJ, Das S. (2009). High multiplication frequency and genetic stability for commercialization of the three varieties of micropropagated tea plants (Camellia spp.). *Scientia Horticulturae, 120*(4), 544-550.

Borse BB. (2012). Novel Bio-Chemical Profiling of Indian Black Teas with Reference to Quality Parameters. *Journal of Microbial & Biochemical Technology, s14*. doi:10.4172/jbb.S14-004

Bothwell JHF, Griffin JL. (2011). An introduction to biological nuclear magnetic resonance spectroscopy. *Biological Reviews, 86*(2), 493-510.

Botten D, Fugallo G, Fraternali F, Molteni C. (2015). Structural Properties of Green Tea Catechins. *J Phys Chem B, 119*(40), 12860-12867. doi:10.1021/acs.jpcb.5b08737

Bowles D, Isayenkova J, Lim E-K, Poppenberger B. (2005). Glycosyltransferases: managers of small molecules. *Current Opinion in Plant Biology, 8*(3), 254-263.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635.

Breseghello F, Sorrells ME. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop science, 46*(3), 1323-1330.

Buratti S, Casiraghi A, Minghetti P, Giovanelli G. (2013). The joint use of electronic nose and electronic tongue for the evaluation of the sensorial properties of green and black tea infusions as related to their chemical composition. *Food and Nutrition Sciences, 4*(6), 605-615.

Caffin N, D'Arcy B, Yao L, Rintoul G. (2004). Developing an index of quality for Australian tea, Rural Industries Research and Development Corporation. In.

Cakir M, Gupta S, Platz GJ, Ablett GA, Loughman R, Emebiri LC, et al. (2003). Mapping and validation of the genes for resistance to Pyrenophora teres f. teres in barley (Hordeum vulgare L.). *Australian Journal of Agricultural Research, 54*(12), 1369-1377.

Carr MKV. (2018). *Advances in tea agronomy*: Cambridge University Press.

Carr MKV, Stephens W. (1992). Climate, weather and the yield of tea. In *Tea* (pp. 87-135): Springer.

Cerrudo D, Cao S, Yuan Y, Martinez C, Suarez EA, Babu R, et al. (2018). Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments. *Frontiers in Plant Science, 9*. doi:10.3389/fpls.2018.00366

Chanda S, De A, Tudu B, Bandyopadhyay R, Hazarika AK, Sabhapondit S, et al. (2016, 21-23 Oct. 2016). *Prediction of polyphenol content in tea leaves using NIR spectroscopy.* Paper presented at the 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI).

Chang HT. (1981). Thea—a section of beveragial tea trees of the genus Camellia. *Acta Sci Nat Univ Sunyats, 1*, 87–99.

Chang Y, Oh EU, Lee MS, Kim HB, Moon D-G, Song KJ. (2017). Construction of a genetic linkage map based on RAPD, AFLP, and SSR markers for tea plant (Camellia sinensis). *Euphytica, 213*(8), 190.

Chaturvedula VSP, Prakash I. (2011). The aroma, taste, color and bioactive constituents of tea. *Journal of Medicinal Plants Research, 5*(11), 2110-2124.

Chen GH, Yang CY, Lee SJ, Wu CC, Tzen JTC. (2014). Catechin content and the degree of its galloylation in oolong tea are inversely correlated with cultivation altitude. *J Food Drug Anal, 22*(3), 303-309. doi:10.1016/j.jfda.2013.12.001

Chen H, Liang H, Ding J, Lai J, Huan Y, Qiao X. (2007a). Rapid Differentiation of Tea Products by Surface Desorption Atmospheric Pressure Chemical Ionization Mass Spectrometry. *Journal of agricultural and food chemistry, 55*(25), 10093-10100. doi:10.1021/jf0720234

Chen L, Wang P-S, Satoshi Y. (2002). Discrimination of wild tea germplasm resources (Camellia sp.) using RAPD markers. *Agricultural Sciences in China, 1*(10), 1105-1110.

Chen L, Yu F, Tong Q. (2000). Discussions on phylogenetic classification and evolution of Sect. Thea. *Journal of Tea Science, 20*(2), 89-94.
Chen L, Zhou Z-X, Yang Y-J. (2007b). Genetic improvement and breeding of tea plant (Camellia sinensis) in China: from individual selection to hybridization and molecular breeding. *Euphytica, 154*(1-2), 239-248.
Chen Q, Chen M, Liu Y, Wu J, Wang X, Ouyang Q, et al. (2018). Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea. *Journal of food science and technology, 55*(10), 4363-4368. doi:10.1007/s13197-018-3353-1
Chen Q, Guo Z, Zhao J. (2008a). Identification of green tea's (Camellia sinensis (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. *Journal of pharmaceutical and biomedical analysis, 48*(5), 1321-1325.
Chen Q, Zhang D, Pan W, Ouyang Q, Li H, Urmila K, et al. (2015). Recent developments of green analytical techniques in analysis of tea's quality and nutrition. *Trends in Food Science & Technology, 43*(1), 63-82.
Chen Q, Zhao J, Cai J. (2008b). Identification of tea varieties using computer vision. *Transactions of the ASABE, 51*(2), 623-628.
Chen Q, Zhao J, Chen Z, Lin H, Zhao D-A. (2011a). Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. *Sensors and actuators B: Chemical, 159*(1), 294-300.
Chen Q, Zhao J, Guo Z, Wang X. (2010). Determination of caffeine content and main catechins contents in green tea (Camellia sinensis L.) using taste sensor technique and multivariate calibration. *Journal of Food Composition and Analysis, 23*(4), 353-358. doi:10.1016/j.jfca.2009.12.010
Chen Q, Zhao J, Liu M, Cai J, Liu J. (2008c). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *Journal of Pharmaceutical and Biomedical Analysis, 46*(3), 568-573.
Chen Q, Zhao J, Vittayapadung S. (2008d). Identification of the green tea grade level using electronic tongue and pattern recognition. *Food Research International, 41*(5), 500-504.
Chen Q, Zhao J, Zhang H, Muhua L, Fang M. (2005). Qualitative Identification of Tea by near Infrared Spectroscopy Based on Soft Independent Modelling of Class Analogy Pattern Recognition. *Journal of Near Infrared Spectroscopy, 13*(6), 327-332.
Chen W-T, Chang H-T. (2013). Tea Identification through Surface-Assisted Laser Desorption/Ionization Mass Spectrometry. *International Journal of Analytical Mass Spectrometry and Chromatography, 1*(01), 11.
Chen X, Snyder CL, Truksa M, Shah S, Weselake RJ. (2011b). sn-Glycerol-3-phosphate acyltransferases in plants. *Plant signaling & behavior, 6*(11), 1695-1699. doi:10.4161/psb.6.11.17777
Chen Z-M, Chen L. (2012). Delicious and healthy tea: an overview. In *Global tea breeding* (pp. 1-11): Springer.
Cheng TO. (2004). Will green tea be even better than black tea to increase coronary flow velocity reserve? *American Journal of Cardiology, 94*(9), 1223.
Chistiakov DA, Hellemans B, Volckaert FAM. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture, 255*(1-4), 1-29.
Ciarkowska A, Ostrowski M, Starzyńska E, Jakubowska A. (2019). Plant SCPL acyltransferases: multiplicity of enzymes with various functions in secondary metabolism. *Phytochemistry Reviews, 18*(1), 303-316.
Ciosek P, Wroblewski W. (2007). Sensor arrays for liquid sensing–electronic tongue systems. *Analyst, 132*(10), 963-978.
Clifford MN, Stoupi S, Kuhnert N. (2007). Profiling and characterization by LC-MS n of the galloylquinic acids of green tea, tara tannin, and tannic acid. *Journal of agricultural and food chemistry, 55*(8), 2797-2807.

Cole M, Covington JA, Gardner JW. (2011). Combined electronic nose and tongue for a flavour sensing system. *Sensors and actuators B: Chemical, 156*(2), 832-839.
Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196.
Collard BCY, Mackill DJ. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1491), 557-572. doi:10.1098/rstb.2007.2170
Collard BCY, Mackill DJ. (2007). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1491), 557-572.
Conesa A, Götz S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics, 2008*.
Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676. doi:10.1093/bioinformatics/bti610
Cordero C, Canale F, Rio DD, Bicchi C. (2009). Identification, quantitation, and method validation for flavan-3-ols in fermented ready-to-drink teas from the Italian market using HPLC-UV/DAD and LC-MS/MS. *Journal of Separation Science, 32*(21), 3643-3651.
Corley RHV, Chomboi KC. (2005). Tea tasting-a statistical evaluation of tasters' skills. *Tea, 26*(1), 10-18.
Corley RHV, Tuwei G. (2018). The Well-Bred Tea Bush. *Cambridge University Press.In: M. Carr (Author), Advances in Tea Agronomy*, 106-136.
Cortina-Puig M, Muñoz-Berbel X, Alonso-Lomillo MA, Muñoz-Pascual FJ, del Valle M. (2007). EIS multianalyte sensing with an automated SIA system—An electronic tongue employing the impedimetric signal. *Talanta, 72*(2), 774-779.
Cros D, Denis M, Sánchez L, Cochard B, Flori A, Durand-Gasselin T, et al. (2014). Genomic selection prediction accuracy in a perennial crop: case study of oil palm (Elaeis guineensis Jacq.). *Theoretical and Applied Genetics, 128*(3), 397-410. doi:10.1007/s00122-014-2439-z
Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3: Genes, Genomes, Genetics*, g3-113.
Crossa J, de Los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics, 186*(2), 713-724.
Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in plant science, 22*(11), 961-975.
Dai J, Patel JD, Mumper RJ. (2007). Characterization of blackberry extract and its antiproliferative and anti-inflammatory properties. *Journal of medicinal food, 10*(2), 258-265.
Dalluge JJ, Nelson BC. (2000). Determination of tea catechins. *Journal of Chromatography A, 881*(1-2), 411-424.
Dana W, Ivo W. (2008). Computer image analysis of seed shape and seed color for flax cultivar description. *Computers and electronics in agriculture, 61*(2), 126-135.
Dare AP, Tomes S, Jones M, McGhie TK, Stevenson DE, Johnson RA, et al. (2013). Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (Malus × domestica). *The Plant Journal, 74*(3), 398-410. doi:10.1111/tpj.12140
Das B, Patel PK, Sobhapandit S, Gogoi RC. (2016). Screening and determination of anthocyanin in pigmented tea germplasm. *Two and a Bud, 63*(1), 1-3.
Das SC, Das, S., , Hazarika M. (2012). Breeding of the Tea Plant (Camellia sinensis) in India. In N. M. (Ed.), *Global Tea Breeding. Advanced Topics in Science and Technology in China.* (pp. 69-124.). Berlin, Heidelberg, : Springer.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics, 12*(7), 499.

Daykin CA, Duynhoven JPMV, Groenewegen A, Dachtler M, Amelsvoort JMMV, Mulder TPJ. (2005). Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. *Journal of agricultural and food chemistry, 53*(5), 1428-1434.

de B. Harrington P, Wang X. (2017). Spectral Representation of Proton NMR Spectroscopy for the Pattern Recognition of Complex Materials. *Journal of Analysis and Testing, 1*(2), 10. doi:10.1007/s41664-017-0003-y

De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG. (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One, 8*(5), e62137.

de Hoffmann E. (2005). Mass Spectrometry. In *Kirk-Othmer Encyclopedia of Chemical Technology*.

de la Torre F, Cañas RA, Pascual MB, Avila C, Cánovas FM. (2014). Plastidic aspartate aminotransferases and the biosynthesis of essential amino acids in plants. *Journal of Experimental Botany, 65*(19), 5527-5534. doi:10.1093/jxb/eru240

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. (2012). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics, 193*(2), 327-345. doi:10.1534/genetics.112.143313

de los Campos G, Pérez-Rodríguez P. (2016). BGLR: Bayesian generalized linear regression. *R package version, 1*(5).

de los Campos G, Sorensen D, Gianola D. (2015). Genomic heritability: what is it? *PLoS genetics, 11*(5), e1005048.

de Roos APW, Hayes BJ, Goddard ME. (2009). Reliability of Genomic Predictions Across Multiple Populations. *Genetics, 183*(4), 1545-1553. doi:10.1534/genetics.109.104935

de Villemereuil P, Morrissey MB, Nakagawa S, Schielzeth H. (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: issues and solutions. *J Evol Biol, 31*(4), 621-632. doi:10.1111/jeb.13232

Deb S, Pou KRJ. (2016). A Review of Withering in the Processing of Black Tea. *Journal of Biosystems Engineering, 41*(4), 365-372.

Del Rio D, Stewart AJ, Mullen W, Burns J, Lean MEJ, Brighenti F, et al. (2004). HPLC-MSn analysis of phenolic compounds and purine alkaloids in green and black tea. *Journal of agricultural and food chemistry, 52*(10), 2807-2815.

Denis M, Bouvet J-M. (2013). Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics & Genomes, 9*(1), 37-51.

Devarumath R, Nandy S, Rani V, Marimuthu S, Muraleedharan N, Raina S. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing Camellia sinensis (China type) and C. assamica ssp. assamica (Assam-India type). *Plant cell reports, 21*(2), 166-173.

Domon B, Aebersold R. (2006). Mass Spectrometry and Protein Analysis. *Science, 312*(5771), 212-217. doi:10.1126/science.1124619

Dong C-w, Zhu H-k, Zhao J-w, Jiang Y-w, Yuan H-b, Chen Q-s. (2017). Sensory quality evaluation for appearance of needle-shaped green tea based on computer vision and nonlinear tools **基于机器视**觉和非线性的芽形绿茶外形感官品质评价. *Journal of Zhejiang University-SCIENCE B, 18*(6), 544-548.

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

El-Dien OG, Ratcliffe B, Klápště J, Chen C, Porth I, El-Kassaby YA. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC genomics, 16*(1), 370.

El-Dien OG, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA. (2018). Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (Picea glauca x engelmannii). *Molecular Breeding, 38*(3), 26.

El-Shahawi MS, Hamza A, Bahaffi SO, Al-Sibaai AA, Abduljabbar TN. (2012a). Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography. *Food chemistry, 134*(4), 2268-2275.

El-Shahawi MS, Hamza A, Bahaffi SO, Al-Sibaai AA, Abduljabbar TN. (2012b). Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography. *Food Chem, 134*(4), 2268-2275. doi:10.1016/j.foodchem.2012.03.039

Ellis R, Nyirenda HE. (1995). A successful plant improvement programme on tea (Camellia sinensis). *Experimental Agriculture, 31*(3), 307-323.

Elosua C, Bariain C, Luquin A, Laguna M, Matias IR. (2012). Optical fiber sensors array to identify beverages by their odor. *IEEE Sensors Journal, 12*(11), 3156-3162.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One, 6*(5), e19379.

Eminoğlu A, Aktürk Dizman Y, Güzel Ş, Beldüz AO. (2018). Molecular and in silico cloning, identification, and preharvest period expression analysis of a putative cytochrome P450 monooxygenase gene from Camellia sinensis (L.) Kuntze (tea). *Turkish journal of biology = Turk biyoloji dergisi, 42*(1), 1-11. doi:10.3906/biy-1606-54

Emwas A-HM. (2015). The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In J. T. Bjerrum (Ed.), *Metabonomics: Methods and Protocols* (pp. 161-193). New York, NY: Springer New York.

Endelman JB. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The plant genome, 4*(3), 250-255.

Escuder-Gilabert L, Peris M. (2010). Highlights in recent applications of electronic tongues in food analysis. *Analytica Chimica Acta, 665*(1), 15-25.

Falcone Ferreyra ML, Rius SP, Casati P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers in Plant Science, 3*, 222-222. doi:10.3389/fpls.2012.00222

Falconer DS, Mackay TFC. (1996). Introduction to quantitative genetics. Longman. *Essex, England*.

Fan F-Y, Shi M, Nie Y, Zhao Y, Ye J-H, Liang Y-R. (2016). Differential behaviors of tea catechins under thermal processing: Formation of non-enzymatic oligomers. *Food Chemistry, 196*, 347-354.

FAO. (2018). FAOSTAT: Download data. Statistics Division, Food and Agriculture Organization of the United Nations,Rome, Italy.

Fikri NA, Adom AH, Ahmad MN, Abdullah AH, Zakaria A, Markom MA. (2011). Development of human sensory mimicking system. *Sensor Letters, 9*(1), 423-427.

Fischer RA, Byerlee DR, Edmeades GO. (2009). *Can technology deliver on the yield challenge to 2050?* Retrieved from

Forgács E, Cserháti T. (2003). CHROMATOGRAPHY | Principles. In B. Caballero (Ed.), *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (pp. 1259-1267). Oxford: Academic Press.

Fraser K, Lane GA, Otter DE, Harrison SJ, Quek SY, Hemar Y, et al. (2013). Monitoring tea fermentation/manufacturing by direct analysis in real time (DART) mass spectrometry. *Food Chem, 141*(3), 2060-2065. doi:10.1016/j.foodchem.2013.05.054

Friedman M, Kim SY, Lee SJ, Han GP, Han JS, Lee KR, et al. (2005). Distribution of catechins, theaflavins, caffeine, and theobromine in 77 teas consumed in the United States. *Journal of food science, 70*(9), C550-C559.

Galkin A, Drose S, Brandt U. (2006). The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. *Biochim Biophys Acta, 1757*(12), 1575-1581. doi:10.1016/j.bbabio.2006.10.001

Gallaher RN, Gallaher K, Marshall AJ, Marshall AC. (2006). Mineral analysis of ten types of commercially available tea. *Journal of Food Composition and Analysis, 19*, S53-S57.

García M, Fernández MJ, Fontecha JL, Lozano J, Santos JP, Aleixandre M, et al. (2006). Differentiation of red wines using an electronic nose based on surface acoustic wave devices. *Talanta, 68*(4), 1162-1165.
Gardner JW, Bartlett PN. (1994). A brief history of electronic noses. *Sensors and actuators B: Chemical, 18*(1-3), 210-211.
Ghosh A, Tamuly P, Bhattacharyya N, Tudu B, Gogoi N, Bandyopadhyay R. (2012). Estimation of theaflavin content in black tea using electronic tongue. *Journal of Food Engineering, 110*(1), 71-79.
Gill GS, Kumar A, Agarwal R. (2011a). Monitoring and grading of tea by computer vision–A review. *Journal of Food Engineering, 106*(1), 13-19.
Gill GS, Kumar A, Agarwal R. (2011b). Monitoring and grading of tea by computer vision – A review. *Journal of Food Engineering, 106*(1), 13-19. doi:https://doi.org/10.1016/j.jfoodeng.2011.04.013
Goddard M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica, 136*(2), 245-257.
Goddard ME, Hayes BJ. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of animal breeding and genetics, 128*(6), 409-421.
Goddard ME, Hayes BJ. (2007). Genomic selection. *Journal of Animal breeding and Genetics, 124*(6), 323-330.
Golding J, Roach P, Parks S. (2009). *Production of high quality export green tea through integrated management*: RIRDC.
Gong J-s, Tang C, Peng C-x. (2012). Characterization of the chemical differences between solvent extracts from Pu-erh tea and Dian Hong black tea by CP–Py–GC/MS. *Journal of analytical and applied pyrolysis, 95*, 189-197.
Goswami H, Barooah AK, Gohain KK, Saikia BP, Baruah A, Dutta A. (2001). Soil parameters for higher productivity of tea. *Two and a Bud, 48*(1), 26-30.
Grattapaglia D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. In *Genomics of plant genetic resources* (pp. 651-682): Springer.
Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, et al. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in plant science, 9*.
Griffiths J. (2008). A brief history of mass spectrometry. *Anal Chem, 80*(15), 5678-5683.
Gu X, Yao C, Zhang Z, Wan X, Ning J, Shao W. (2011). GC–ECD method for determination of glucosidically bound aroma precursors in fresh tea leaves. *Chromatographia, 73*(1-2), 189-193.
Gunasekare MTK. (2012). Tea plant (Camellia sinensis) breeding in Sri Lanka. In *Global Tea Breeding* (pp. 125-176): Springer.
Gunathilaka RPD, Smart JCR, Fleming CM. (2017). The impact of changing climate on perennial crops: the case of tea production in Sri Lanka. *Climatic change, 140*(3-4), 577-592.
Gupta PK, Rustgi S, Kulwal PL. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant molecular biology, 57*(4), 461-485.
Gupta PK, Varshney RK, Sharma PC, Ramesh B. (1999). Molecular markers and their applications in wheat breeding. *Plant breeding, 118*(5), 369-390.
Guth H, Grosch W. (1993). Identification of potent odourants in static headspace samples of green and black tea powders on the basis of aroma extract dilution analysis (AEDA). *Flavour and fragrance journal, 8*(4), 173-178.
Habier D, Fernando RL, Dekkers JCM. (2009). Genomic selection using low-density marker panels. *Genetics*.
Habier D, Fernando RL, Kizilkaya K, Garrick DJ. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics, 12*(1), 186.
Hackett CA, Wachira FN, Paul S, Powell W, Waugh R. (2000). Construction of a genetic linkage map for Camellia sinensis (tea). *Heredity, 85*(4), 346.
Haddi Z, Mabrouk S, Bougrini M, Tahri K, Sghaier K, Barhoumi H, et al. (2014). E-Nose and e-Tongue combination for improved recognition of fruit juice samples. *Food Chemistry, 150*, 246-253.

Hall D, Tegström C, Ingvarsson PK. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in Functional Genomics, 9*(2), 157-165.
Han W-Y, Huang J-G, Li X, Li Z-X, Ahammed GJ, Yan P, et al. (2017). Altitudinal effects on the quality of green tea in east China: a climate change perspective. *European Food Research and Technology, 243*(2), 323-330.
Han Z-X, Rana MM, Liu G-F, Gao M-J, Li D-X, Wu F-G, et al. (2016). Green tea flavour determinants and their changes over manufacturing processes. *Food Chemistry, 212*, 739-748.
Hara Y. (2001). *Green tea: health benefits and applications*: CRC press.
Hara Y, Luo SJ, Wickremashinghe RL, Yamanishi T. (1995). IV. Biochemistry of processing black tea. *Food Reviews International, 11*, 457-471.
Harbowy ME, Balentine DA, Davies AP, Cai Y. (1997). Tea chemistry. *Critical reviews in plant sciences, 16*(5), 415-480.
Hayashi N, Chen R, Hiraoka M, Ujihara T, Ikezaki H. (2010). β-Cyclodextrin/surface plasmon resonance detection system for sensing bitter-astringent taste intensity of green tea catechins. *Journal of Agricultural and Food Chemistry, 58*(14), 8351-8356.
Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. (2009a). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution, 41*(1), 51.
Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. (2009b). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science, 92*(2), 433-443. doi:10.3168/jds.2008-1646
Hayes BJ, Cogan NOI, Pembleton LW, Goddard ME, Wang J, Spangenberg GC, et al. (2013). Prospects for genomic selection in forage plant species. *Plant breeding, 132*(2), 133-143.
Hayes BJ, Goddard ME. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.
Hayes BJ, Visscher PM, Goddard ME. (2009c). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research, 91*(1), 47-60.
Hazarika K. (2012). Domestic Consumer Market for Indian Tea: A Survey Report. *International Journal of Latest Trends in Finance and Economic Sciences, 2*(3), 251-256.
He X, Li J, Zhao W, Liu R, Zhang L, Kong X. (2015). Chemical fingerprint analysis for quality control and identification of Ziyang green tea by HPLC. *Food chemistry, 171*, 405-411.
Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop science, 51*(6), 2597-2606.
Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Science, 50*(5), 1681-1690.
Heffner EL, Sorrells ME, Jannink J-L. (2009a). Genomic Selection for Crop Improvement. *Crop Science, 49*(1). doi:10.2135/cropsci2008.08.0512
Heffner EL, Sorrells ME, Jannink J-L. (2009b). Genomic selection for crop improvement. *Crop Science, 49*(1), 1-12.
Henderson CR. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447.
Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, et al. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop science, 54*(4), 1476-1488.
Hindorff LA. (2009). A catalog of published genome-wide association studies. *http://www. genome. gov/26525384*.
Ho C-T, Zheng X, Li S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9-27.
Hodgson JM, Croft KD. (2010). Tea flavonoids and cardiovascular health. *Molecular aspects of medicine, 31*(6), 495-502.
Holland JB. (2004, 2004). *Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities*.

Hong X, Wang J. (2014). Detection of adulteration in cherry tomato juices based on electronic nose and tongue: Comparison of different data fusion approaches. *Journal of Food Engineering, 126*, 89-97.
Horie H, Kohata K. (1998). Application of capillary electrophoresis to tea quality estimation. *Journal of Chromatography A, 802*(1), 219-223.
Hou YJ, He Q, Liang GL, Li PW, Peng P, Deng M. (2006). ISSR analysis of the hybrid of the descendants of tea camellias. *Journal of Southwest Agricultural University (Natural Science), 28*(2), 267-270.
Hsieh M-M, Chen S-M. (2007). Determination of amino acids in tea leaves and beverages using capillary electrophoresis with light-emitting diode-induced fluorescence detection. *Talanta, 73*(2), 326-331.
Huang JA, Li JX, Huang YH, Luo JW, Gong ZH, Liu ZH. (2005). Construction of AFLP molecular markers linkage map in tea plant. *J Tea Sci, 25*(1), 7-15.
Huang Y-b, Lan Y-b, Lacey RE. (2004). Artificial senses for characterization of food quality. *Journal of Bionic Engineering, 1*(4), 159-173. doi:10.1007/bf03399472
Huang Z, Zhang Y, Wang L, Ding L, Wang M, Yan H, et al. (2009). Simultaneous determination of 103 pesticide residues in tea samples by LC-MS/MS. *Journal of separation science, 32*(9), 1294-1301.
ITC. (2018). International Tea Committee-Annual Bulletin of Statistics. . *International Tea Committee.*
.
ITC. (2017). Annual Bulletin of Statistics. . from International Tea Committee.

Ivarsson P, Holmin S, Höjer N-E, Krantz-Rülcker C, Winquist F. (2001). Discrimination of tea by means of a voltammetric electronic tongue and different applied waveforms. *Sensors and actuators B: Chemical, 76*(1-3), 449-454.
Jaccoud D, Peng K, Feinstein D, Kilian A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic acids research, 29*(4), e25-e25.
Jain SM, Priyadarshan PM. (2009). *Breeding plantation tree crops: tropical species*: Springer.
James D, Scott SM, Ali Z, O'hare WT. (2005). Chemical sensors for electronic nose systems. *Microchimica Acta, 149*(1-2), 1-17.
Jan HU, Abbadi A, Lücke S, Nichols RA, Snowdon RJ. (2016). Genomic prediction of testcross performance in canola (Brassica napus). *PLoS One, 11*(1), e0147769.
Jannink J-L, Lorenz AJ, Iwata H. (2010a). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics, 9*(2), 166-177.
Jannink JL, Lorenz AJ, Iwata H. (2010b). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics, 9*(2), 166-177. doi:10.1093/bfgp/elq001
Jeganathan B, Kottawa-Arachchi JD, Ranatunga MAB, Abeysinghe ISB, Gunasekare MTK, Bandara BMR. (2017). Genotypic variation in biochemical compounds of the Sri Lankan tea (Camellia sinensis L.) accessions and their relationships to quality and biotic stresses AU - Punyasiri, P. A. Nimal. *The Journal of Horticultural Science and Biotechnology, 92*(5), 502-512. doi:10.1080/14620316.2017.1289070
Jeong B, Song Y, Moon Y, Han S, Bang J, Kim J, et al. (2005, Nov 11–15.). *Tea tree breeding plans for the tea Industry in Korea.* Paper presented at the International Tea Symposium-2005, China.
Jiang G-L. (2013). Molecular markers and marker-assisted breeding in plants. *Plant breeding from laboratories to fields*, 45-83.
Jiang H, Engelhardt UH, Thräne C, Maiwald B, Stark J. (2015). Determination of flavonol glycosides in green tea, oolong tea and black tea by UHPLC compared to HPLC. *Food chemistry, 183*, 30-35.
Jin J-Q, Yao M-Z, Ma C-L, Ma J-Q, Chen L. (2016). Association mapping of caffeine content with TCS1 in tea plant and its related species. *Plant physiology and biochemistry, 105*, 251-259.
John KMM, Thiruvengadam M, Enkhtaivan G, Kim DH. (2014). Variation in major phenolic compounds and quality potential of CTC black tea elicited by Saccharomyces cercevisiae and its correlation with antioxidant potential. *Industrial Crops and Products, 55*, 289-294.
Jones P, Vogt T. (2001). Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta, 213*(2), 164-174.

Juliana P, Singh RP, Singh PK, Crossa J, Huerta-Espino J, Lan C, et al. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theoretical and Applied Genetics, 130*(7), 1415-1430.

Kamunya SM, Wachira FN, Pathak RS, Korir R, Sharma V, Kumar R, et al. (2010). Genomic mapping and testing for quantitative trait loci in tea (Camellia sinensis (L.) O. Kuntze). *Tree genetics & genomes, 6*(6), 915-929.

Kamunya SM, Wachira FN, Pathak RS, Muoki RC, Sharma RK. (2012). Tea improvement in Kenya. In *Global Tea Breeding* (pp. 177-226): Springer.

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

Kaundun SS, Zhyvoloup A, Park Y-G. (2000). Evaluation of the genetic diversity among elite tea (Camellia sinensis var. sinensis) accessions using RAPD markers. *Euphytica, 115*(1), 7-16.

Kawakami M. (1997). Comparison of extraction techniques for characterizing tea aroma and analysis of tea by GC-FTIR-MS. In *Plant Volatile Analysis* (pp. 211-229): Springer.

Kawakami M, Ganguly SN, Banerjee J, Kobayashi A. (1995). Aroma composition of oolong tea and black tea by brewed extraction method and characterizing compounds of Darjeeling tea aroma. *Journal of Agricultural and Food Chemistry, 43*(1), 200-207.

Kawakami M, Sarma S, Himizu K, Konishi Y, Kobayashi A. (2004, 2004). *Aroma characteristics of Darjeeling tea*.

Kerio LC, Wachira FN, Wanyoko JK, Rotich MK. (2012). Characterization of anthocyanins in Kenyan teas: Extraction and identification. *Food chemistry, 131*(1), 31-38. doi:https://doi.org/10.1016/j.foodchem.2011.08.005

Khokhar S, Venema D, Hollman PCH, Dekker M, Jongen W. (1997). A RP-HPLC method for the determination of tea catechins. *Cancer letters, 114*(1-2), 171-172.

Kiani S, Minaei S, Ghasemi-Varnamkhasti M. (2016). Fusion of artificial senses as a robust approach to food quality assessment. *Journal of Food Engineering, 171*, 230-239.

Kiehne A, Engelhardt UH. (1996). Thermospray-LC-MS analysis of various groups of polyphenols in tea. *Zeitschrift für Lebensmittel-Untersuchung und Forschung, 202*(1), 48-54.

Kiehne A, Lakenbrink C, Engelhardt UH. (1997). Analysis of proanthocyanidins in tea samples I. LC-MS results. *Zeitschrift für Lebensmitteluntersuchung und-Forschung A, 205*(2), 153-157.

Kim KS, Park SH, Choung MG. (2007). Nondestructive determination of oil content and fatty acid composition in perilla seeds by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 55*(5), 1679-1685.

Koch KR. (1990). Quantitative determination of aluminium in tea by means of aluminium-27 nuclear magnetic resonance spectroscopy. *Analyst, 115*(6), 823-825.

Koch W, Kukula-Koch W, Komsta L, Marzec Z, Szwerc W, Glowniak K. (2018). Green Tea Quality Evaluation Based on Its Catechins and Metals Composition in Combination with Chemometric Analysis. *Molecules, 23*(7). doi:10.3390/molecules23071689

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Apostolides Z. (2018a). Identification of novel QTL for black tea quality traits and drought tolerance in tea plants (Camellia sinensis). *Tree genetics & genomes, 14*(1), 9.

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Apostolides Z. (2018b). Identification of novel QTL for black tea quality traits and drought tolerance in tea plants (Camellia sinensis). *Tree Genetics & Genomes, 14*(1). doi:10.1007/s11295-017-1219-8

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Joubert F, et al. (2019). Functional annotation of putative QTL associated with black tea quality and drought tolerance traits. *Scientific Reports, 9*(1), 1465. doi:10.1038/s41598-018-37688-z

Kondo K. (1977). Chromosome numbers in the genus Camellia. *Biotropica*, 86-94.

Kondo K, Parks CR. (1981). Cytological studies in cultivated species of Camellia. VI. Giemsa C-banded karyotypes of seven accessions of Camellia japonica L. sensu lato. *Japanese Journal of Breeding, 31*(1), 26-34.

Korte A, Farlow A. (2013a). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods, 9*(1), 29.
Korte A, Farlow A. (2013b). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods, 9*, 29. doi:10.1186/1746-4811-9-29
Kruijer W, Boer MP, Malosetti M, Flood PJ, Engel B, Kooke R, et al. (2015). Marker-based estimation of heritability in immortal populations. *Genetics, 199*(2), 379-398.
Kumamoto M, Sonda T, Takedomi K, Tabata M. (2000). Enhanced separation and elution of catechins in HPLC using mixed-solvents of water, acetonitrile and ethyl acetate as the mobile phase. *Analytical Sciences, 16*(2), 139-144.
Kumar D, Gulati A, Sharma U. (2015). Determination of Theanine and Catechin in Camellia sinensis (Kangra Tea) Leaves by HPTLC and NMR Techniques. *Food Analytical Methods, 9*(6), 1666-1674. doi:10.1007/s12161-015-0343-z
Kumar PV, Basheer S, Ravi R, Thakur MS. (2011a). Comparative assessment of tea quality by various analytical and sensory methods with emphasis on tea polyphenols. *J Food Sci Technol, 48*(4), 440-446. doi:10.1007/s13197-010-0178-y
Kumar PVS, Basheer S, Ravi R, Thakur MS. (2011b). Comparative assessment of tea quality by various analytical and sensory methods with emphasis on tea polyphenols. *Journal of food science and technology, 48*(4), 440-446.
Kumar RSS, Murugesan S, Kottur G, Gyamfi D. (2013). Black tea: The plants, processing/manufacturing and production. In *Tea in health and disease prevention* (pp. 41-57): Elsevier.
Kumazawa K, Masuda H. (2002). Identification of potent odorants in different green tea varieties using flavor dilution technique. *Journal of Agricultural and Food Chemistry, 50*(20), 5660-5663.
Kumazawa S, Kajiya K, Naito A, SaitÔ H, Tuzi S, Tanio M, et al. (2004). Direct Evidence of Interaction of a Green Tea Polyphenol, Epigallocatechin Gallate, with Lipid Bilayers by Solid-state Nuclear Magnetic Resonance. *Bioscience, biotechnology, and biochemistry, 68*(8), 1743-1747. doi:10.1271/bbb.68.1743
Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, et al. (2017). Genomic Selection in Commercial Perennial Crops: Applicability and Improvement in Oil Palm (Elaeis guineensis Jacq.). *Sci Rep, 7*(1), 2872. doi:10.1038/s41598-017-02602-6
Lairson LL, Henrissat B, Davies GJ, Withers SG. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annual review of biochemistry, 77*.
Larive CK, Barding GA, Dinges MM. (2015). NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Analytical Chemistry, 87*(1), 133-146. doi:10.1021/ac504075g
Lawless HT, Heymann H. (2010). *Sensory evaluation of food: principles and practices*: Springer Science & Business Media.
Le Gall G, Colquhoun IJ, Defernez M. (2004). Metabolite profiling using 1H NMR spectroscopy for quality assessment of green tea, Camellia sinensis (L.). *Journal of Agricultural and Food Chemistry, 52*(4), 692-700.
Lee B-L, Ong C-N. (2000). Comparative analysis of tea catechins and theaflavins by high-performance liquid chromatography and capillary electrophoresis. *Journal of Chromatography A, 881*(1-2), 439-447.
Lee J-E, Lee B-J, Chung J-O, Hwang J-A, Lee S-J, Lee C-H, et al. (2010). Geographical and climatic dependencies of green tea (Camellia sinensis) metabolites: a 1H NMR-based metabolomics study. *Journal of agricultural and food chemistry, 58*(19), 10582-10589.
Lee J-E, Lee B-J, Chung J-O, Shin H-J, Lee S-J, Lee C-H, et al. (2011). 1H NMR-based metabolomic characterization during green tea (Camellia sinensis) fermentation. *Food Research International, 44*(2), 597-604. doi:https://doi.org/10.1016/j.foodres.2010.12.004
Lee J, Chambers D, Chambers E, Adhikari K, Yoon Y. (2013). Volatile Aroma Compounds in Various Brewed Green Teas. *Molecules, 18*(8), 10024-10041. doi:10.3390/molecules180810024
Lee J, Chambers DH. (2007). A lexicon for flavor descriptive analysis of green tea. *Journal of Sensory Studies, 22*(3), 256-272.

Lee LS, Kim SH, Kim YB, Kim YC. (2014a). Quantitative analysis of major constituents in green tea with different plucking periods and their antioxidant activity. *Molecules, 19*(7), 9173-9186. doi:10.3390/molecules19079173

Lee MS, Hwang YS, Lee J, Choung MG. (2014b). The characterization of caffeine and nine individual catechins in the leaves of green tea (Camellia sinensis L.) by near-infrared reflectance spectroscopy. *Food Chem, 158*, 351-357. doi:10.1016/j.foodchem.2014.02.127

Leng P-f, Lübberstedt T, Xu M-l. (2017). Genomics-assisted breeding – A revolutionary strategy for crop improvement. *Journal of Integrative Agriculture, 16*(12), 2674-2685. doi:10.1016/s2095-3119(17)61813-6

Li L, Xie S, Ning J, Chen Q, Zhang Z. (2018). Evaluating green tea quality based on multisensor data fusion combining hyperspectral imaging and olfactory visualization systems. *Journal of the Science of Food and Agriculture, 0*(0). doi:doi:10.1002/jsfa.9371

Li M, Zhou J, Gu X, Wang Y, Huang X, Yan C. (2009). Quantitative capillary electrophoresis and its application in analysis of alkaloids in tea, coffee, coca cola, and theophylline tablets. *Journal of Separation Science, 32*(2), 267-274.

Li S, Lo CY, Pan MH, Lai CS, Ho CT. (2013). Black tea: chemical analysis and stability. *Food Funct, 4*(1), 10-18. doi:10.1039/c2fo30093a

Liang Y, Lu J, Zhang L, Wu S, Wu Y. (2003). Estimation of black tea quality by analysis of chemical composition and colour difference of tea infusions. *Food chemistry, 80*(2), 283-290.

Liang YR, Ye Q, Jin J, Liang H, Lu JL, Du YY, et al. (2008). Chemical and instrumental assessment of green tea sensory preference. *International Journal of Food Properties, 11*(2), 258-272.

Lin J, Dai Y, Guo Y-n, Xu H-r, Wang X-c. (2012). Volatile profile analysis and quality prediction of Longjing tea (Camellia sinensis) by HS-SPME/GC-MS. *Journal of Zhejiang University Science B, 13*(12), 972-980.

Lin L-Z, Chen P, Harnly JM. (2008). New phenolic components and chromatographic profiles of green and fermented teas. *Journal of Agricultural and Food Chemistry, 56*(17), 8130-8140.

Lin X. (2009). Qualitative and quantitative analysis method of tea by near infrared spectroscopy. *Spectroscopy and Spectral Analysis, 29*(9), 2417-2420.

Lipka AE, Lu F, Cherney JH, Buckler ES, Casler MD, Costich DE. (2014). Accelerating the switchgrass (Panicum virgatum L.) breeding cycle using genomic selection approaches. *PloS one, 9*(11), e112227.

Liu M, Tian H-l, Wu J-H, Cang R-R, Wang R-X, Qi X-H, et al. (2015). Relationship between gene expression and the accumulation of catechin during spring and autumn in tea plants (Camellia sinensis L.). *Horticulture research, 2*(1), 15011. doi:10.1038/hortres.2015.11

Liu Z, Xie H-l, Chen L, Huang J-h. (2018). An Improved Weighted Partial Least Squares Method Coupled with Near Infrared Spectroscopy for Rapid Determination of Multiple Components and Anti-Oxidant Activity of Pu-Erh Tea. *Molecules, 23*(5), 1058.

Liu ZJ, Cordes JF. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1-4), 1-37.

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. (2011a). Genomic Selection in Plant Breeding. In (pp. 77-123).

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. (2011b). Genomic selection in plant breeding: knowledge and prospects. In *Advances in agronomy* (Vol. 110, pp. 77-123): Elsevier.

Lorenzana RE, Bernardo R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and applied genetics, 120*(1), 151-161.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS genetics, 9*(1), e1003215.

Lubang'a NM, Kamunya SM, Kiplagat O, Wanyoko JK, Chalo RM. (2015). Effects of genotype by environment interaction on quality traits in Kenyan tea (Camellia sinensis)(L.) O. Kuntze. *Tea, 36*(1), 24-39.

Lv H-P, Zhong Q-S, Lin Z, Wang L, Tan J-F, Guo L. (2012). Aroma characterisation of Pu-erh tea using headspace-solid phase microextraction combined with GC/MS and GC–olfactometry. *Food chemistry, 130*(4), 1074-1081.
Lv Y, Yang X, Zhao Y, Ruan Y, Yang Y, Wang Z. (2009). Separation and quantification of component monosaccharides of the tea polysaccharides from Gynostemma pentaphyllum by HPLC with indirect UV detection. *Food Chemistry, 112*(3), 742-746.
Lvova L, Legin A, Vlasov Y, Cha GS, Nam H. (2003). Multicomponent analysis of Korean green tea by means of disposable all-solid-state potentiometric electronic tongue microsystem. *Sensors and actuators B: Chemical, 95*(1-3), 391-399.
Ma J-Q, Huang L, Ma C-L, Jin J-Q, Li C-F, Wang R-K, et al. (2015). Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PloS one, 10*(6), e0128798.
Ma J-Q, Yao M-Z, Ma C-L, Wang X-C, Jin J-Q, Wang X-M, et al. (2014). Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (Camellia sinensis). *PloS one, 9*(3), e93131.
Ma X, Feng F, Wei H, Mei H, Xu K, Chen S, et al. (2016). Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. *Frontiers in Plant Science, 7*, 1801.
Mackay IJ, Mackay IJ, Caligari PDS, Gibson JP. (1999). Accelerated recurrent selection. *Euphytica, 105*(1), 43-51. doi:10.1023/a:1003428430664
Mackay TFC. (2001). The genetic architecture of quantitative traits. *Annual review of genetics, 35*(1), 303-339.
Magagna F, Cordero C, Cagliero C, Liberto E, Rubiolo P, Sgorbini B, et al. (2017). Black tea volatiles fingerprinting by comprehensive two-dimensional gas chromatography - Mass spectrometry combined with high concentration capacity sample preparation techniques: Toward a fully automated sensomic assessment. *Food Chem, 225*, 276-287. doi:10.1016/j.foodchem.2017.01.003
Magoma GN, Wachira FN, Obanda M, Imbuga M, Agong SG. (2000). The use of catechins as biochemical markers in diversity studies of tea (Camellia sinensis). *Genetic Resources and Crop Evolution, 47*(2), 107-114.
Mahindroo N. (2000). Tea-Camellia sinensis-report. *ICS-UNIDO, Trieste, Italy*.
Mahrous EA, Farag MA. (2015). Two dimensional NMR spectroscopic approaches for exploring plant metabolome: a review. *Journal of advanced research, 6*(1), 3-15.
Manley M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews, 43*(24), 8200-8214.
Marchini J, Howie B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics, 11*(7), 499.
Martens S, Preuß A, Matern U. (2010). Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in Arabidopsis thaliana L. *Phytochemistry, 71*(10), 1040-1049.
Marx W, Haunschild R, Bornmann L. (2017). Global warming and tea production—The bibliometric view on a newly emerging research topic. *Climate, 5*(3), 46.
Matsumoto S, Kiriiwa Y, Takeda Y. (2002). Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA. *Theoretical and applied genetics, 104*(6-7), 998-1002.
Matsumoto S, Kiriiwa Y, Yamaguchi S. (2004). The Korean tea plant (Camellia sinensis): RFLP analysis of genetic diversity and relationship to Japanese tea. *Breeding science, 54*(3), 231-237.
Mazza G. (2007). Anthocyanins and heart health. *Annali-Istituto Superiore Di Sanita, 43*(4), 369.
McCouch SR, Zhao K, Wright M, Tung C-W, Ebana K, Thomson M, et al. (2010). Development of genome-wide SNP assays for rice. *Breeding science, 60*(5), 524-535.
Meegahakumbura MK, Wambulwa MC, Li DZ, Gao LM. (2018a). Preliminary Investigations on the Genetic Relationships and Origin of Domestication of the Tea Plant (Camellia sinensis (L.)) Using Genotyping by Sequencing. *Tropical Agricultural Research, 29*(3), 229 – 240.

Meegahakumbura MK, Wambulwa MC, Li M-M, Thapa KK, Sun Y-S, Möller M, et al. (2018b). Domestication origin and breeding history of the tea plant (Camellia sinensis) in China and India based on nuclear microsatellites and cpDNA sequence data. *Frontiers in plant science, 8*, 2270.

Meegahakumbura MK, Wambulwa MC, Thapa KK, Li MM, Möller M, Xu JC, et al. (2016). Indications for three independent domestication events for the tea plant (Camellia sinensis (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PloS one, 11*(5), e0155369.

Melorose J, Perroy R, Careas S. (2015). World population prospects. *United Nations, 1*(6042), 587-592.

Menet M-C, Sang S, Yang CS, Ho C-T, Rosen RT. (2004). Analysis of theaflavins and thearubigins from black tea extract by MALDI-TOF mass spectrometry. *Journal of agricultural and food chemistry, 52*(9), 2455-2461.

Meng W, Xu X, Cheng K-K, Xu J, Shen G, Wu Z, et al. (2017). Geographical Origin Discrimination of Oolong Tea (TieGuanYin, Camellia sinensis (L.) O. Kuntze) Using Proton Nuclear Magnetic Resonance Spectroscopy and Near-Infrared Spectroscopy. *Food Analytical Methods, 10*(11), 3508-3522. doi:10.1007/s12161-017-0920-4

Meuwissen THE, Hayes BJ, Goddard ME. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.

Meuwissen THE, Hayes, B. J., Goddard, M. E. . (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829.

Meuwissen THE, Solberg TR, Shepherd R, Woolliams JA. (2009). A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genetics Selection Evolution, 41*(1), 2.

Ming TL. (2000). Monograph of the Genus Camellia. Yunnan Science and Technology Press. *Kunming, China*, 128-134.

Mishra P, Nordon A, Tschannerl J, Lian G, Redfern S, Marshall S. (2018a). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering, 238*, 70-77. doi:https://doi.org/10.1016/j.jfoodeng.2018.06.015

Mishra P, Nordon A, Tschannerl J, Lian G, Redfern S, Marshall S. (2018b). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering*.

Mishra RK, Sen-Mandi S. (2004). Molecular profiling and development of DNA marker associated with drought tolerance in tea clones growing in Darjeeling. *Current Science*, 60-66.

Mochida K, Yamazaki Y, Ogihara Y. (2004). Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetics and Genomics, 270*(5), 371-377.

Mondal TK. (2014a). *Breeding and biotechnology of tea and its wild species*: Springer Science & Business Media.

Mondal TK. (2014b). Genetics and Breeding. In *Breeding and Biotechnology of Tea and its Wild Species* (pp. 9-34).

Mondal TK. (2011). Camellia. In *Wild crop relatives: genomic and breeding resources* (pp. 15-39): Springer.

Mondal TK, Bhattacharya A, Laxmikumaran M, Ahuja PS. (2004). Recent advances of tea (Camellia sinensis) biotechnology. *Plant Cell, Tissue and Organ Culture, 76*(3), 195-254.

Moreno L, Merlos A, Abramova N, Jimenez C, Bratov A. (2006). Multi-sensor array used as an "electronic tongue" for mineral water analysis. *Sensors and actuators B: Chemical, 116*(1-2), 130-134.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genetics Selection Evolution, 41*(1), 56.

Mphangwe NIK, Vorster J, Steyn JM, Nyirenda HE, Taylor NJ, Apostolides Z. (2013). Screening of tea (Camellia sinensis) for trait-associated molecular markers. *Applied biochemistry and biotechnology, 171*(2), 437-449.

Msomba SW, Kamunya SM, Rweyemamu CL, Reuben SOWM. (2018). EFFECT OF STABILITY AND ADAPTABILITY ON TEA QUALITY IN DIVERSE GROWING ENVIRONMENTS OF TANZANIA.

Mukhopadhyay M, Mondal TK, Chand PK. (2016). Biotechnological advances in tea (Camellia sinensis [L.] O. Kuntze): a review. *Plant cell reports, 35*(2), 255-287.

Mukhtar H, Ahmad N. (2000a). Tea polyphenols: prevention of cancer and optimizing health. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.

Mukhtar H, Ahmad N. (2000b). Tea polyphenols: prevention of cancer and optimizing health–. *The American journal of clinical nutrition, 71*(6), 1698S-1702S.

Mulder HA, Calus MPL, Druet T, Schrooten C. (2012). Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *Journal of Dairy Science, 95*(2), 876-889.

Müller BSF, de Almeida Filho JE, Lima BM, Garcia CC, Missiaggia A, Aguiar AM, et al. (2019). Independent and Joint-GWAS for growth traits in Eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytologist, 221*(2), 818-833.

Muoki RC, Wachira FN, Pathak RS, Kamunya SM. (2007). Assessment of the mating system of Camellia sinensis in biclonal seed orchards based on PCR markers. *The Journal of Horticultural Science and Biotechnology, 82*(5), 733-738.

Muranty H, Troggio M, Sadok IB, Al Rifaï M, Auwerkerken A, Banchi E, et al. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture research, 2*, 15060.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, et al. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell, 21*(8), 2194-2202.

Nakabayashi T. (1994). Chemical components in tea leaves. *Chemistry and functions of green, black and Oolong teas*, 20-42.

Namal Senanayake SPJ. (2013). Green tea extract: Chemistry, antioxidant properties and food applications – A review. *Journal of Functional Foods, 5*(4), 1529-1541. doi:10.1016/j.jff.2013.08.011

Namita P, Mukesh R, Vijay KJ. (2012). Camellia sinensis (green tea): A review. *Global journal of pharmacology, 6*(2), 52-59.

Narukawa M, Kimata H, Noga C, Watanabe T. (2010). Taste characterisation of green tea catechins. *International journal of food science & technology, 45*(8), 1579-1585.

Narukawa M, Noga C, Ueno Y, Sato T, Misaka T, Watanabe T. (2011). Evaluation of the bitterness of green tea catechins by a cell-based assay with the human bitter taste receptor hTAS2R39. *Biochemical and biophysical research communications, 405*(4), 620-625.

Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, et al. (2004). Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome research, 14*(8), 1664-1668.

Niessen WMA. (2001). Principles and instrumentation of gas chromatography-mass spectrometry. In *Current Practice of Gas Chromatography-Mass Spectrometry* (pp. 15-43): CRC Press.

Ning P, Gong C, Zhang Y, Guo K. (2010). La, Ce, Pr, Nd and Sm concentrations in Pu'er tea of Yunnan, China. *Journal of Rare Earths(English Edition), 28*(4), 636-640.

Novak I, Šeruga M, Komorsky-Lovrić Š. (2010). Characterisation of catechins in green and black teas using square-wave voltammetry and RP-HPLC-ECD. *Food Chemistry, 122*(4), 1283-1289.

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. *Journal of plant physiology, 198*, 39-48.

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

Obanda M, Owuor PO, Mang'oka R, Kavoi MM. (2004). Changes in thearubigin fractions and theaflavin levels due to variations in processing conditions and their influence on black tea liquor brightness and total colour. *Food Chemistry, 85*(2), 163-173.

Obanda M, Owuor PO, Taylor SJ. (1997). Flavanol composition and caffeine content of green leaf as quality potential indicators of Kenyan black teas. *Journal of the Science of Food and Agriculture, 74*(2), 209-215.

Okinda Owuor P, Obanda M, Nyirenda HE, Mphangwe NIK, Wright LP, Apostolides Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (Camellia sinensis) produced in Kenya and comparison with similar teas from Malawi and South Africa. *Food Chemistry, 97*(4), 644-653. doi:10.1016/j.foodchem.2005.04.027

Olafsdottir G, Martinsdóttir E, Oehlenschläger J, Dalgaard P, Jensen B, Undeland I, et al. (1997). Methods to evaluate fish freshness in research and industry. *Trends in Food Science & Technology, 8*(8), 258-265.

Osborne BG. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation*.

Owuor PO. (1992). Comparison of gas chromatographic volatile profiling methods for assessing the flavour quality of Kenyan black teas. *Journal of the Science of Food and Agriculture, 59*(2), 189-197.

Owuor PO, Kamau DM, Kamunya SM, Msomba SW, Uwimana MA, Okal AW, et al. (2011). Effects of genotype, environment and management on yields and quality of black tea. In *Genetics, Biofuels and Local Farming Systems* (pp. 277-307): Springer.

Owuor PO, Mang'oka R, Kavoi MM. (2004). Changes in thearubigin fractions and theaflavin levels due to variations in processing conditions and their influence on black tea liquor brightness and total colour.

Owuor PO, McDowell I. (1994). Changes in theaflavin composition and astringency during black tea fermentation. *Food Chemistry, 51*(3), 251-254. doi:10.1016/0308-8146(94)90023-x

Owuor PO, Obanda M. (2017). Clonal Variations in the Response of Hard Physically Withered Leaf to Rehydration Following Long Chemical Wither Durations.

Owuor PO, Obanda M. (2007). The use of green tea (Camellia sinensis) leaf flavan-3-ol composition in predicting plain black tea quality potential. *Food chemistry, 100*(3), 873-884.

Owuor PO, Obanda M. (1998). The changes in black tea quality due to variations of plucking standard and fermentation time. *Food Chemistry, 61*(4), 435-441.

Owuor PO, Obanda M. (1997). Advances in withering technologies and future strategies to cope with high tea production in Kenya. *Tea, 18*(2), 184-193.

Owuor PO, Obanda M, Nyirenda HE, Mphangwe NIK, Wright LP, Apostolides Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (Camellia sinensis) produced in Kenya and comparison with similar teas from Malawi and South Africa. *Food Chemistry, 97*(4), 644-653.

Owuor PO, Obanda MA, Othieno CO, Horita H, Tsushida T, Murai T. (1987). Changes in the chemical composition and quality of black tea due to plucking standards. *Agricultural and Biological Chemistry, 51*(12), 3383-3384.

Owuor PO, Tsushida T, Horita H, Murai T. (1988). Effects of geographical area of production on the composition of the volatile flavour compounds in Kenyan clonal black CTC teas. *Experimental Agriculture, 24*(2), 227-235.

Palit M, Tudu B, Dutta PK, Dutta A, Jana A, Roy JK, et al. (2010). Classification of black tea taste and correlation with tea taster's mark using voltammetric electronic tongue. *IEEE transactions on instrumentation and measurement, 59*(8), 2230-2239.

Panigrahi N, Bhol CS, Das BS. (2016). Rapid assessment of black tea quality using diffuse reflectance spectroscopy. *Journal of Food Engineering, 190*, 101-108.

Parra V, Arrieta ÁA, Fernández-Escudero JA, García H, Apetrei C, Rodríguez-Méndez ML, et al. (2006). E-tongue based on a hybrid array of voltammetric sensors based on phthalocyanines, perylene derivatives and conducting polymers: Discrimination capability towards red wines elaborated with different varieties of grapes. *Sensors and actuators B: Chemical, 115*(1), 54-61.

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC plant biology, 12*(1), 16.

Paul S, Wachira FN, Powell W, Waugh R. (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (Camellia sinensis (L.) O. Kuntze) revealed by AFLP markers. *Theoretical and applied genetics, 94*(2), 255-263.

Pavlou AK, Magan N, Jones JM, Brown J, Klatser P, Turner APF. (2004). Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system. *Biosensors and Bioelectronics, 20*(3), 538-544.

Payne RW. (2015). *Introduction to GenStat® for windows™*. Retrieved from Hemel Hempstead:

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM. (2009). GenStat for windows introduction. *VSN International, Hemel Hempstead, 204*.

Pelillo M, Bonoli M, Biguzzi B, Bendini A, Toschi TG, Lercker G. (2004). An investigation in the use of HPLC with UV and MS-electrospray detection for the quantification of tea catechins. *Food Chemistry, 87*(3), 465-470.

Peters CM, Green RJ, Janle EM, Ferruzzi MG. (2010). Formulation with ascorbic acid and sucrose modulates catechin bioavailability from green tea. *Food Research International, 43*(1), 95-102.

Pongsuwan W, Bamba T, Harada K, Yonetani T, Kobayashi A, Fukusaki E. (2008a). High-throughput technique for comprehensive analysis of Japanese green tea quality assessment using ultra-performance liquid chromatography with time-of-flight mass spectrometry (UPLC/TOF MS). *Journal of agricultural and food chemistry, 56*(22), 10705-10708.

Pongsuwan W, Bamba T, Yonetani T, Kobayashi A, Fukusaki E. (2008b). Quality prediction of Japanese green tea using pyrolyzer coupled GC/MS based metabolic fingerprinting. *Journal of agricultural and food chemistry, 56*(3), 744-750.

Poon GK. (1998). Analysis of catechins in tea extracts by liquid chromatography–electrospray ionization mass spectrometry. *Journal of Chromatography A, 794*(1-2), 63-74.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904.

Pripdeevech P, Machan T. (2011). Fingerprint of volatile flavour constituents and antioxidant activities of teas from Thailand. *Food Chemistry, 125*(2), 797-802.

Punyasiri PAN, Jeganathan B, Kottawa-Arachchi JD, Ranatunga MAB, Abeysinghe ISB, Gunasekare MTK, et al. (2017). Genotypic variation in biochemical compounds of the Sri Lankan tea (Camellia sinensis L.) accessions and their relationships to quality and biotic stresses. *The Journal of Horticultural Science and Biotechnology, 92*(5), 502-512.

R. Core T. (2015). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. R Foundation for Statistical Computing. In.

Rafalski JA. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science, 162*(3), 329-333.

Raina SN, Ahuja PS, Sharma RK, Das SC, Bhardwaj P, Negi R, et al. (2012). Genetic structure and diversity of India hybrid tea. *Genetic Resources and Crop Evolution, 59*(7), 1527-1541.

Rakhmawati A, Yuliastuti E. (2011). *Compact computer vision for black tea quality evaluation based on the black tea particles*.

Ravichandran R. (2002). Carotenoid composition, distribution and degradation to flavour volatiles during black tea manufacture and the effect of carotenoid supplementation on tea quality and aroma. *Food chemistry, 78*(1), 23-28.

Ravichandran R, Parthiban R. (1998). The impact of processing techniques on tea volatiles. *Food chemistry, 62*(3), 347-353.

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

Ren G, Wang S, Ning J, Xu R, Wang Y, Xing Z, et al. (2013). Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *Food Research International, 53*(2), 822-826.

Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, et al. (2012a). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist, 194*(1), 116-128. doi:10.1111/j.1469-8137.2011.04038.x

Resende MFR, Munoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, et al. (2012b). Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (Pinus taeda L.). *Genetics, 190*(4), 1503-1510. doi:10.1534/genetics.111.137026

Reto M, Figueira ME, Filipe HM, Almeida CMM. (2007). Analysis of vitamin K in green tea leafs and infusions by SPME–GC-FID. *Food Chemistry, 100*(1), 405-411.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics, 44*(2), 217.

Röck F, Barsan N, Weimar U. (2008). Electronic nose: current status and future trends. *Chemical Reviews, 108*(2), 705-725.

Rodríguez-Méndez ML, De Saja JA, González-Antón R, García-Hernández C, Medina-Plaza C, García-Cabezón C, et al. (2016). Electronic Noses and Tongues in Wine Industry. *Frontiers in Bioengineering and Biotechnology, 4*. doi:10.3389/fbioe.2016.00081

Roy RB, Modak A, Mondal S, Tudu B, Bandyopadhyay R, Bhattacharyya N. (2013). Fusion of electronic nose and tongue response using fuzzy based approach for black tea classification. *Procedia Technology, 10*, 615-622.

Roy RB, Tudu B, Shaw L, Jana A, Bhattacharyya N, Bandyopadhyay R. (2012). Instrumental testing of tea by combining the responses of electronic nose and tongue. *Journal of Food Engineering, 110*(3), 356-363.

Roy SC, Chakraborty BN. (2009). Genetic diversity and relationships among tea (Camellia sinensis) cultivars as revealed by RAPD and ISSR based fingerprinting.

Rutkoski JE, Poland J, Jannink J-L, Sorrells ME. (2013). Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes, Genomes, Genetics, 3*(3), 427-439.

Samanta T, Cheeni V, Das S, Roy AB, Ghosh BC, Mitra A. (2015). Assessing biochemical changes during standardization of fermentation time and temperature for manufacturing quality black tea. *J Food Sci Technol, 52*(4), 2387-2393. doi:10.1007/s13197-013-1230-5

Samynathan R, Perisamy Sk, Gandhi S, Anitha J, Sanmugam G, Padmanabhan M, et al. (2016). Biochemical and molecular analysis of Camellia sinensis (L.) O. Kuntze tea from the selected P/11/15 clone. *Journal of Genetic Engineering and Biotechnology, 14*(1), 69-75. doi:https://doi.org/10.1016/j.jgeb.2015.12.004

Sang S, Tian S, Stark RE, Yang CS, Ho C-T. (2004). New dibenzotropolone derivatives characterized from black tea using LC/MS/MS. *Bioorganic & medicinal chemistry, 12*(11), 3009-3017.

Sano M, Tabata M, Suzuki M, Degawa M, Miyase T, Maeda-Yamamoto M. (2001). Simultaneous determination of twelve tea catechins by high-performance liquid chromatography with electrochemical detection. *Analyst, 126*(6), 816-820.

Sapozhnikova Y. (2014). Development of liquid chromatography–tandem mass spectrometry method for analysis of polyphenolic compounds in liquid samples of grape juice, green tea and coffee. *Food Chemistry, 150*, 87-93.

Scampicchio M, Benedetti S, Brunetti B, Mannino S. (2006). Amperometric electronic tongue for the evaluation of the tea astringency. *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis, 18*(17), 1643-1648.

Scharbert S, Hofmann T. (2005). Molecular definition of black tea taste by means of quantitative studies, taste reconstitution, and omission experiments. *Journal of Agricultural and Food Chemistry, 53*(13), 5377-5384.

Scharbert S, Holzmann N, Hofmann T. (2004). Identification of the astringent taste compounds in black tea infusions by combining instrumental analysis and human bioresponse. *Journal of agricultural and food chemistry, 52*(11), 3498-3508.
Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics, 167*(1), 485-498.
Schuh C, Schieberle P. (2006). Characterization of the key aroma compounds in the beverage prepared from Darjeeling black tea: quantitative differences between tea leaves and infusion. *Journal of agricultural and food chemistry, 54*(3), 916-924.
Schuler MA. (1996). Plant cytochrome P450 monooxygenases. *Critical reviews in plant sciences, 15*(3), 235-284.
Sehra G, Cole M, Gardner JW. (2004). Miniature taste sensing system based on dual SH-SAW sensor device: an electronic tongue. *Sensors and actuators B: Chemical, 103*(1-2), 233-239.
Semagn K, Bjørnstad Å, Ndjiondjop MN. (2006a). An overview of molecular marker methods for plants. *African journal of biotechnology, 5*(25).
Semagn K, Bjørnstad Å, Ndjiondjop MN. (2006b). Principles, requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology, 5*(25).
Semagn K, Bjørnstad Å, Xu Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology, 13*(5), 16-17.
Seneca. (2007). CHAPTER 2 - Alkaloid Chemistry. In T. Aniszewski (Ed.), *Alkaloids - Secrets of Life* (pp. 61-139). Amsterdam: Elsevier.
Sharangi AB. (2009). Medicinal and therapeutic potentialities of tea (Camellia sinensis L.)–A review. *Food Research International, 42*(5-6), 529-535.
Sharma M, Thomas EV. (2013). Electronic vision study of tea grains color during Infusion. *International Journal of Engineering Science Invention, 2*(4), 52-58.
Sharma P, Ghosh A, Tudu B, Sabhapondit S, Baruah BD, Tamuly P, et al. (2015). Monitoring the fermentation process of black tea using QCM sensor based electronic nose. *Sensors and actuators B: Chemical, 219*, 146-157.
Sharma RK, Negi MS, Sharma S, Bhardwaj P, Kumar R, Bhattachrya E, et al. (2010). AFLP-based genetic diversity assessment of commercially important tea germplasm in India. *Biochemical genetics, 48*(7-8), 549-564.
Shellie R, Marriott P, Cornwell C. (2000). Characterization and comparison of tea tree and lavender oils by using comprehensive gas chromatography. *Journal of High Resolution Chromatography, 23*(9), 554-560.
Sinija VR, Mishra HN. (2011). Fuzzy analysis of sensory data for quality evaluation and ranking of instant green tea powder and granules. *Food and Bioprocess Technology, 4*(3), 408-416.
Sitienei BJ, Juma SG, Opere E. (2017). On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County, Kenya. *Climate, 5*(3), 54.
Skov T, Bro R. (2005). A new approach for modelling sensor based data. *Sensors and actuators B: Chemical, 106*(2), 719-729.
Sliwinska M, Wisniewska P, Dymerski T, Namiesnik J, Wardencki W. (2014). Food analysis using artificial senses. *Journal of Agricultural and Food Chemistry, 62*(7), 1423-1448.
Smyth H, Cozzolino D. (2012). Instrumental Methods (Spectroscopy, Electronic Nose, and Tongue) As Tools To Predict Taste and Aroma in Beverages: Advantages and Limitations. *Chemical Reviews, 113*(3), 1429-1440. doi:10.1021/cr300076c
Snel J, Lorist MM. (2011). Effects of caffeine on sleep and cognition. In *Progress in brain research* (Vol. 190, pp. 105-117): Elsevier.
Sohn Y-S, Goodey A, Anslyn EV, McDevitt JT, Shear JB, Neikirk DP. (2005). A microbead array chemical sensor using capillary-based sample introduction: toward the development of an "electronic tongue". *Biosensors and Bioelectronics, 21*(2), 303-312.

Someswararao C, Srivastav PP, Das H. (2013). Quality of black teas in Indian market. *African Journal of Agricultural Research, 8*(5), 491-494.
Song J, Carver BF, Powers C, Yan L, Klápště J, El-Kassaby YA, et al. (2017). Practical application of genomic selection in a doubled-haploid winter wheat breeding program. *Molecular Breeding, 37*(10). doi:10.1007/s11032-017-0715-8
Sorrells ME, Yu J. (2009). Linkage disequilibrium and association mapping in the Triticeae. In *Genetics and genomics of the Triticeae* (pp. 655-683): Springer.
Soto-Cerda BJ, Cloutier S. (2012a). Association mapping in plant genomes. *Genetic Diversity in Plants*, 29-54.
Soto-Cerda BJ, Cloutier S. (2012b). Association mapping in plant genomes. In *Genetic diversity in plants*: InTech.
Spindel J, Iwata H. (2018). Genomic Selection in Rice Breeding. In T. Sasaki, M. Ashikari (Eds.), *Rice Genomics, Genetics and Breeding* (pp. 473-496). Singapore: Springer Singapore.
Steinmann D, Ganzera M. (2011). Recent advances on HPLC/MS in medicinal plant analysis. *Journal of Pharmaceutical and Biomedical Analysis, 55*(4), 744-757.
Stich B, Melchinger AE. (2010). An introduction to association mapping in plants. *CAB Rev, 5*, 1-9.
Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. (2005). Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theoretical and applied genetics, 111*(4), 723-730.
Stich B, Van Inghelandt D. (2018). Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. *Frontiers in Plant Science, 9*, 159-159. doi:10.3389/fpls.2018.00159
Stone H, Sidel JL. (2004). Sensory Evaluation Practices, Elsevier Academic Press. *California, USA*.
Strike DJ, Meijerink MGH, Koudelka-Hep M. (1999). Electronic noses–A mini-review. *Fresenius' journal of analytical chemistry, 364*(6), 499-505.
Su H, Liu J, He W, Wen B, Li M, Feng J. (2016). Association analysis of traits related to tea (Camellia sinensis) quality with EST-SSRs in Southern Henan area. *Journal of Agricultural Biotechnology, 24*(9), 1328-1336.
Su J, Zhang F, Chong X, Song A, Guan Z, Fang W, et al. (2019). Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. *Horticulture Research, 6*(1), 21. doi:10.1038/s41438-018-0101-7
Sultana T, Stecher G, Mayer R, Trojer L, Qureshi MN, Abel G, et al. (2008). Quality assessment and quantitative analysis of flavonoids from tea samples of different origins by HPLC-DAD-ESI-MS. *Journal of agricultural and food chemistry, 56*(10), 3444-3453.
Sun D-W. (2016). *Computer vision technology for food quality evaluation*: Academic Press.
Suzuki Y, Shioi Y. (2003). Identification of chlorophylls and carotenoids in major teas by high-performance liquid chromatography with photodiode array detection. *Journal of agricultural and food chemistry, 51*(18), 5307-5314.
Sverrisdóttir E, Byrne S, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, et al. (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics, 130*(10), 2091-2108.
Tahara Y, Toko K. (2013). Electronic tongues–a review. *IEEE Sensors Journal, 13*(8), 3001-3011.
Takeo T. (1992). Green and semi-fermented teas. In *Tea* (pp. 413-457): Springer.
Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F 1 hybrids. *BMC Plant Biology, 17*(1), 110.
Tan L-Q, Wang L-Y, Xu L-Y, Wu L-Y, Peng M, Zhang C-C, et al. (2016a). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (Camellia sinensis). *Tree genetics & genomes, 12*(3), 52.

Tan L-Q, Wang L-Y, Xu L-Y, Wu L-Y, Peng M, Zhang C-C, et al. (2016b). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (Camellia sinensis). *Tree Genetics & Genomes, 12*(3), 52. doi:10.1007/s11295-016-1008-9

Tanaka J. (2012). Japanese Tea Breeding History and the Future Perspective. In *Global Tea Breeding: Achievements, Challenges and Perspectives* (pp. 227-239). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tanaka T, Mizutani T, Shibata M, Tanikawa N, Parks CR. (2005a). Cytogenetic studies on the origin of Camellia× vernalis. V. Estimation of the seed parent of C.× vernalis that evolved about 400 years ago by cpDNA analysis. *Journal of the Japanese Society for Horticultural Science, 74*(6), 464-468.

Tanaka T, Mizutani T, Shibata M, Tanikawa, N., , Parks CR. (2005b). Cytogenetic studies on the origin of Camellia × vernalis. V. Estimation of the seed parent of C. × vernalis that evolved about 400 years ago by cpDNA analysis. . *J Jap Soc Hort Sci., 74*, 464–468.

Tang P, Shen D-Y, Xu Y-Q, Zhang X-C, Shi J, Yin J-F. (2018). Effect of Fermentation Conditions and Plucking Standards of Tea Leaves on the Chemical Components and Sensory Quality of Fermented Juice. *Journal of Chemistry, 2018*, 1-7. doi:10.1155/2018/4312875

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, et al. (2012). Construction of a high-density reference linkage map of tea (Camellia sinensis). *Breeding science, 62*(3), 263-273.

Taniguchi F, Kimura K, Saba T, Ogino A, Yamaguchi S, Tanaka J. (2014). Worldwide core collections of tea (Camellia sinensis) based on SSR markers. *Tree genetics & genomes, 10*(6), 1555-1565.

Tao W, Zhou Z, Zhao B, Wei T. (2016a). Simultaneous determination of eight catechins and four theaflavins in green, black and oolong tea using new HPLC-MS-MS method. *J Pharm Biomed Anal, 131*, 140-145. doi:10.1016/j.jpba.2016.08.020

Tao W, Zhou Z, Zhao B, Wei T. (2016b). Simultaneous determination of eight catechins and four theaflavins in green, black and oolong tea using new HPLC–MS–MS method. *Journal of pharmaceutical and biomedical analysis, 131*, 140-145.

Tarachiwin L, Ute K, Kobayashi A, Fukusaki E. (2007). 1H NMR based metabolic profiling in the evaluation of Japanese green tea quality. *Journal of agricultural and food chemistry, 55*(23), 9330-9336.

TBK. (2018). Tea industry performance in 2018. *The Tea Board of Kenya.*

TBK. (2017). Tea industry performance in 2017. from The Tea Board of Kenya.

Thakur J, Gupta R. (2012). Improvement of tea leaves fermentation through pectinases. *Acta Microbiol Immunol Hung, 59*(3), 321-334. doi:10.1556/AMicr.59.2012.3.3

Toan NV, Tao NV. (2005, Nov 11–15). *Tea breeding selection by hybridization method in Vietnam.* Paper presented at the International Tea Symposium 2005, Hangzhou, China.

Togari N, Kobayashi A, Aishima T. (1995). Pattern recognition applied to gas chromatographic profiles of volatile components in three tea categories. *Food Research International, 28*(5), 495-502.

Tozlu BH, Okumuş Hİ. (2018). A new approach to automation of black tea fermentation process with electronic nose. *Automatika, 59*(3-4), 373-381.

Tu Y, Bian M, Wan Y, Fei T. (2018). Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. *PeerJ, 6*, e4858.

Tudu B, Jana A, Metla A, Ghosh D, Bhattacharyya N, Bandyopadhyay R. (2009). Electronic nose for black tea quality evaluation by an incremental RBF network. *Sensors and actuators B: Chemical, 138*(1), 90-95.

Ujihara T, Ohta R, Hayashi N, Kohata K, Tanaka J-i. (2009). Identification of Japanese and Chinese green tea cultivars by using simple sequence repeat markers to encourage proper labeling. *Bioscience, biotechnology, and biochemistry, 73*(1), 15-20.

Usai MG, Goddard ME, Hayes BJ. (2009). LASSO with cross-validation for genomic selection. *Genetics research, 91*(6), 427-436.

Valin H, Sands RD, Van der Mensbrugghe D, Nelson GC, Ahammad H, Blanc E, et al. (2014). The future of food demand: understanding differences in global economic models. *Agricultural Economics, 45*(1), 51-67.

Valsangiacomo C, Baggi F, Gaia V, Balmelli T, Peduzzi R, Piffaretti J-C. (1995). Use of amplified fragment length polymorphism in molecular typing of Legionella pneumophila and application to epidemiological studies. *Journal of Clinical Microbiology, 33*(7), 1716-1719.

van der Hooft JJJ, Akermi M, Ünlü FY, Mihaleva V, Roldan VG, Bino RJ, et al. (2012). Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts. *Journal of agricultural and food chemistry, 60*(36), 8841-8850.

Van der Vossen HAM, Umali BE. (2000). *Plant resources of South-East Asia*: Backhuys Publ.

VanRaden PM. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science, 91*(11), 4414-4423.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science, 92*(1), 16-24.

Varshney RK, Terauchi R, McCouch SR. (2014). Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLOS Biology, 12*(6), e1001883. doi:10.1371/journal.pbio.1001883

Vasefi F, MacKinnon N, Farkas DL. (2016). Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology. In M. R. Hamblin, P. Avci, G. K. Gupta (Eds.), *Imaging in Dermatology* (pp. 187-201). Boston: Academic Press.

Venzie JL, Castro J, Krishna MVB, Nelson DM, Marcus RK. (2007). Electron-impact and glow-discharge ionization LC–MS analysis of green tea tincture. *Analytical and bioanalytical chemistry, 387*(1), 321-333.

Viskelis P, Rubinskienė M, Jasutienė I, Šarkinas A, Daubaras R, Česonienė L. (2009). Anthocyanins, antioxidative, and antimicrobial properties of American cranberry (Vaccinium macrocarpon Ait.) and their press cakes. *Journal of food science, 74*(2), C157-C161.

Vo-Dinh T. (2004). A hyperspectral imaging system for in vivo optical diagnostics. *IEEE Engineering in Medicine and Biology Magazine, 23*(5), 40-49.

Von Gadow A, Joubert E, Hansmann CF. (1997). Comparison of the antioxidant activity of rooibos tea (Aspalathus linearis) with green, oolong and black tea. *Food chemistry, 60*(1), 73-77.

Wachira F, Tanaka J, Takeda Y. (2001). Genetic variation and differentiation in tea (Camellia sinensis) germplasm revealed by RAPD and AFLP variation. *The Journal of Horticultural Science and Biotechnology, 76*(5), 557-563.

Wachira FN. (2002). Genetic diversity and characterisation of Kenyan tea germplasm. *A tea Genetic Diversity (TGD) project. TGD final project document, Kericho, Kenya*.

Wachira FN, Kamunya SK. (2005). Pseudo-self-incompatibility in some tea clones (Camellia sinensis (L.) O. Kuntze). *The Journal of Horticultural Science and Biotechnology, 80*(6), 716-720. doi:10.1080/14620316.2005.11512004

Wachira FN, Kiplangat JK. (1991). Newly identified Kenyan Polyploid Tea Strains. . *Tea, 12*(1), 10-13.

Wachira FN, Waugh R, Powell W, Hackett CA. (1995). Detection of genetic diversity in tea (Camellia sinensis) using RAPD markers. *Genome, 38*(2), 201-210.

Wambulwa MC, Meegahakumbura MK, Chalo R, Kamunya S, Muchugi A, Xu JC, et al. (2016). Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree genetics & genomes, 12*(1), 11.

Wang D, Lu J, Miao A, Xie Z, Yang D. (2008a). HPLC-DAD-ESI-MS/MS analysis of polyphenols and purine alkaloids in leaves of 22 tea cultivars in China. *Journal of Food Composition and Analysis, 21*(5), 361-369.

Wang H, Helliwell K, You X. (2000a). Isocratic elution system for the determination of catechins, caffeine and gallic acid in green tea using HPLC. *Food chemistry, 68*(1), 115-121.

Wang H, Provan GJ, Helliwell K. (2000b). Tea flavonoids: their functions, utilisation and analysis. *Trends in Food Science & Technology, 11*(4), 152-160. doi:https://doi.org/10.1016/S0924-2244(00)00061-3

Wang J, Wang Y, Cheng J, Wang J, Sun X, Sun S, et al. (2018a). Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy. *Lwt, 96*, 90-97. doi:10.1016/j.lwt.2018.05.012

Wang K, Liu F, Liu Z, Huang J, Xu Z, Li Y, et al. (2010). Analysis of chemical components in oolong tea in relation to perceived quality. *International Journal of Food Science & Technology, 45*(5), 913-920. doi:10.1111/j.1365-2621.2010.02224.x

Wang K, Ruan J. (2009). Analysis of chemical components in green tea in relation with perceived quality, a case study with Longjing teas. *International Journal of Food Science & Technology, 44*(12), 2476-2484.

Wang L-F, Lee J-Y, Chung J-O, Baik J-H, So S, Park S-K. (2008b). Discrimination of teas with different degrees of fermentation by SPME–GC analysis of the characteristic volatile flavour compounds. *Food Chemistry, 109*(1), 196-206.

Wang L-S, Stoner GD. (2008c). Anthocyanins and their role in cancer prevention. *Cancer letters, 269*(2), 281-290.

Wang X, Xu Y, Hu Z, Xu C. (2018b). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal, 6*(4), 330-340. doi:https://doi.org/10.1016/j.cj.2018.03.001

Wang X, Yang Z, Xu C. (2015). A comparison of genomic selection methods for breeding value prediction. *Science Bulletin, 60*(10), 925-935. doi:https://doi.org/10.1007/s11434-015-0791-2

Wang Y, Gao L, Shan Y, Liu Y, Tian Y, Xia T. (2012). Influence of shade on flavonoid biosynthesis in tea (Camellia sinensis (L.) O. Kuntze). *Scientia Horticulturae, 141*, 7-16.

Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, et al. (2018a). Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proceedings of the National Academy of Sciences, 115*(18), E4151-E4158.

Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, et al. (2018b). Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proceedings of the National Academy of Sciences*, 201719622.

Wei K, Wang L-Y, Zhou J, He W, Zeng J-M, Jiang Y-W, et al. (2012). Comparison of catechins and purine alkaloids in albino and normal green tea cultivars (Camellia sinensis L.) by HPLC. *Food chemistry, 130*(3), 720-724.

Wei K, Wang L, Zhou J, He W, Zeng J, Jiang Y, et al. (2011). Catechin contents in tea (Camellia sinensis) as affected by cultivar and environment and their relation to chlorophyll contents. *Food Chemistry, 125*(1), 44-48.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, et al. (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences, 101*(26), 9915-9920.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research, 18*(22), 6531-6535.

Willson KC, Clifford MN. (2012). *Tea: Cultivation to consumption*: Springer Science & Business Media.

Winter P, Kahl G. (1995). Molecular marker technologies for plant improvement. *World Journal of Microbiology and Biotechnology, 11*(4), 438-448.

Wolfe MD, Del Carpio DP, Alabi O, Ezenwaka LC, Ikeogu UN, Kayondo IS, et al. (2017). Prospects for Genomic Selection in Cassava Breeding. *The Plant Genome, 10*(3). doi:10.3835/plantgenome2017.03.0015

Wright LP. (2002). Biochemical analysis for identification of quality in black tea (Camellia sinensis).

Wright LP, Aucamp JP, Apostolides Z. (2001). Analysis of black tea theaflavins by non-aqueous capillary electrophoresis. *Journal of Chromatography A, 919*(1), 205-213.

Wright LP, Mphangwe NIK, Nyirenda HE, Apostolides Z. (2002). Analysis of the theaflavin composition in black tea (Camellia sinensis) for predicting the quality of tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 82*(5), 517-525.

Wright LP, Mphangwe NIK, Nyirenda HE, Apostolides Z. (2000a). Analysis of caffeine and flavan-3-ol composition in the fresh leaf of Camellia sinesis for predicting the quality of the black tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 80*(13), 1823-1830. doi:doi:10.1002/1097-0010(200010)80:13<1823::AID-JSFA702>3.0.CO;2-E

Wright LP, Mphangwe NIK, Nyirenda HE, Apostolides Z. (2000b). Analysis of caffeine and flavan-3-ol composition in the fresh leaf of Camellia sinesis for predicting the quality of the black tea produced in Central and Southern Africa. *Journal of the Science of Food and Agriculture, 80*(13), 1823-1830.

Wrolstad RE, Durst RW, Lee J. (2005). Tracking color and pigment changes in anthocyanin products. *Trends in Food Science & Technology, 16*(9), 423-428.

Wu F, He Z, Luo Q. (1998). High-performance liquid chromatographic determination of oxalic acid in tea using tris (1, 10-phenanthroline)-ruthenium (II) chemiluminescence. *Analytical Sciences, 14*(5), 971-973.

Wu J, Liu J, Fu M, Li G. (2006). Classification of Chinese green tea by a voltammetric electronic tongue. *Chinese Journal of Sensors and Actuators, 4*, 007.

Wu LJ, Peng SZ, Xian Li J, Wen SC, Huan HY, Hua GZ, et al. (2002). Studies on genetic diversity of tea cultivars [Camellia sinensis (L.) O. Kuntze] by RAPD analysis. *J Tea Sci, 12*, 121-127.

Wu R, Zhao J, Chen Q, Huang X. (2011). Quality assessment of green tea taste by using electronic tongue. *Transactions of the Chinese Society of Agricultural Engineering, 27*(11), 378-381.

Wu Z-J, Ma H-Y, Zhuang J. (2018). iTRAQ-based proteomics monitors the withering dynamics in postharvest leaves of tea plant (Camellia sinensis). *Molecular Genetics and Genomics, 293*(1), 45-59.

Xia E-H, Tong W, Wu Q, Wei S, Zhao J, Zhang Z-Z, et al. (2020). Tea plant genomics: achievements, challenges and perspectives. *Horticulture research, 7*(1), 7. doi:10.1038/s41438-019-0225-4

Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, et al. (2017a). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Molecular plant, 10*(6), 866-877.

Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, et al. (2017b). The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Molecular Plant, 10*(6), 866-877. doi:10.1016/j.molp.2017.04.002

Xia EH, Li FD, Tong W, Li PH, Wu Q, Zhao HJ, et al. (2019). Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol J*. doi:10.1111/pbi.13111

Xia EH, Zhang HB, Sheng J, Li K, Zhang QJ, Kim C, et al. (2017c). The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Mol Plant, 10*(6), 866-877. doi:10.1016/j.molp.2017.04.002

Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M, et al. (2005). DArT for high-throughput genotyping of cassava (Manihot esculenta) and its wild relatives. *Theoretical and applied genetics, 110*(6), 1092-1098.

Xiao H, Wang J. (2009). Discrimination of Xihulongjing tea grade using an electronic tongue. *African Journal of Biotechnology, 8*(24).

Xie H, Wang Y, Ding Y, Qiu C, Sun L, Gai Z, et al. (2019). Global Ubiquitome Profiling Revealed the Roles of Ubiquitinated Proteins in Metabolic Pathways of Tea Leaves in Responding to Drought Stress. *Scientific Reports, 9*(1), 4286-4286. doi:10.1038/s41598-019-41041-3

Xiong C, Liu C, Pan W, Ma F, Xiong C, Qi L, et al. (2015). Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system. *Food Chemistry, 176*, 130-136.

Xu L, Deng D-H, Cai C-B. (2011). Predicting the Age and Type of Tuocha Tea by Fourier Transform Infrared Spectroscopy and Chemometric Data Analysis. *Journal of Agricultural and Food Chemistry, 59*(19), 10461-10469. doi:10.1021/jf2026499

Xu LY, Wang LY, Wei K, Tan LQ, Su JJ, Cheng H. (2018a). High-density SNP linkage map construction and QTL mapping for flavonoid-related traits in a tea plant (Camellia sinensis) using 2b-RAD sequencing. *BMC Genomics, 19*(1), 955. doi:10.1186/s12864-018-5291-8

Xu N, Chen ZM. (2002). Green tea, black tea and semi-fermented tea. *Tea: Bioactivity and Therapeutic Potential*, 35-57.

Xu S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics, 163*(2), 789-801.

Xu Y-Q, Wang C, Li C-W, Liu S-H, Zhang C-X, Li L-W, et al. (2016). Characterization of aroma-active compounds of pu-erh tea by headspace solid-phase microextraction (HS-SPME) and simultaneous distillation-extraction (SDE) coupled with GC-olfactometry and GC-MS. *Food Analytical Methods, 9*(5), 1188-1198.

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

Xu Y. (2010a). *Molecular plant breeding*: Cabi.

Xu Y, Li P, Yang Z, Xu C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal, 5*(2), 175-184.

Xu YQ. (2010b). Molecular plant breeding. *CABI, Wallingford, UK*, 736

Yamamoto T, Juneja LR, Chu DC, Kim M. (1997). Chemistry and Applications of Green Tea CRC press. *Florida, USA*.

Yamanishi T. (1995). Special issue on tea: Flavor of tea. *Food Review Int., 11*, 477-525.

Yamauchi Y, Nakamura A, Kohno I, Kitai M, Hatanaka K, Tanimoto T. (2008). Simple and rapid UV spectrophotometry of caffeine in tea coupled with sample pre-treatment using a cartridge column filled with polyvinylpolypyrrolidone (PVPP). *Chemical and Pharmaceutical Bulletin, 56*(2), 185-188.

Yang H, Wei C-L, Liu H-W, Wu J-L, Li Z-G, Zhang L, et al. (2016). Genetic divergence between Camellia sinensis and its wild relatives revealed via genome-wide SNPs from RAD sequencing. *PloS one, 11*(3), e0151424.

Yang T-L, Hao R-H, Guo Y, Papasian CJ, Deng H-W. (2018a). Chapter 4 - Copy Number Variation. In R. V. Thakker, M. P. Whyte, J. A. Eisman, T. Igarashi (Eds.), *Genetics of Bone Biology and Skeletal Disease (Second Edition)* (pp. 43-54): Academic Press.

Yang XR, Ye CX, Xu JK, Jiang YM. (2007). Simultaneous analysis of purine alkaloids and catechins in Camellia sinensis, Camellia ptilophylla and Camellia assamica var. kucha by HPLC. *Food Chemistry, 100*(3), 1132-1136.

Yang Y-Q, Yin H-X, Yuan H-B, Jiang Y-W, Dong C-W, Deng Y-L. (2018b). Characterization of the volatile components in green tea by IRAE-HS-SPME/GC-MS combined with multivariate analysis. *PLoS One, 13*(3), e0193393.

Yang Z, Baldermann S, Watanabe N. (2013). Recent studies of the volatile compounds in tea. *Food Research International, 53*(2), 585-599.

Yao M-Z, Chen L. (2012a). Tea Germplasm and Breeding in China. In *Global Tea Breeding: Achievements, Challenges and Perspectives* (pp. 13-68). Berlin, Heidelberg: Springer Berlin Heidelberg.

Yao M-Z, Chen L. (2012b). Tea germplasm and breeding in China. In *Global Tea Breeding* (pp. 13-68): Springer.

Yaroshenko I, Kirsanov D, Kartsova L, Bhattacharyya N, Sarkar S, Legin A. (2014). On the application of simple matrix methods for electronic tongue data processing: Case study with black tea samples. *Sensors and actuators B: Chemical, 191*, 67-74.

Yashin AY, Nemzer BV, Combet E, Yashin YI. (2015). Determination of the chemical composition of tea by chromatographic methods: a review. *Journal of Food Research, 4*(3), 56-87.

Ye J, Wang W, Ho C, Li J, Guo X, Zhao M, et al. (2016). Differentiation of two types of pu-erh teas by using an electronic nose and ultrasound-assisted extraction-dispersive liquid–liquid microextraction-gas chromatography-mass spectrometry. *Analytical Methods, 8*(3), 593-604.

Ye N, Zhang L, Gu X. (2012). Discrimination of green teas from different geographical origins by using HS-SPME/GC–MS and pattern recognition methods. *Food Analytical Methods, 5*(4), 856-860.

Ye Y, Yan J, Cui J, Mao S, Li M, Liao X, et al. (2018). Dynamic changes in amino acids, catechins, caffeine and gallic acid in green tea during withering. *Journal of Food Composition and Analysis, 66*, 98-108. doi:10.1016/j.jfca.2017.12.008

Yu H, Wang J. (2007). Discrimination of LongJing green-tea grade by electronic nose. *Sensors and actuators B: Chemical, 122*(1), 134-140.

Yu H, Wang J, Xiao H, Liu M. (2009). Quality grade identification of green tea using the eigenvalues of PCA based on the E-nose signals. *Sensors and actuators B: Chemical, 140*(2), 378-382.

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 Science and Technology, 41*(7), 1268-1273.

Yu J, Buckler ES. (2006a). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology, 17*(2), 155-160. doi:10.1016/j.copbio.2006.02.003

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. (2006b). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics, 38*(2), 203.

Yu P, Yeo AS, Low MY, Zhou W. (2014). Identifying key non-volatile compounds in ready-to-drink green tea and their impact on taste profile. *Food Chem, 155*, 9-16. doi:10.1016/j.foodchem.2014.01.046

Yu X, Bai G, Luo N, Chen Z, Liu S, Liu J, et al. (2011). Association of simple sequence repeat (SSR) markers with submergence tolerance in diverse populations of perennial ryegrass. *Plant Science, 180*(2), 391-398. doi:https://doi.org/10.1016/j.plantsci.2010.10.013

Yuan Y, Song Y, Jing W, Wang Y, Yang X, Liu D. (2014). Simultaneous determination of caffeine, gallic acid, theanine,(−)-epigallocatechin and (−)-epigallocatechin-3-gallate in green tea using quantitative 1 H-NMR spectroscopy. *Analytical Methods, 6*(3), 907-914.

Yue J, Zhu C, Zhou Y, Niu X, Miao M, Tang X, et al. (2018). Transcriptome analysis of differentially expressed unigenes involved in flavonoid biosynthesis during flower development of Chrysanthemum morifolium 'Chuju'. *Scientific Reports, 8*(1), 13414.

Zampetti E, Pantalei S, Macagnano A, Proietti E, Di Natale C, D'amico A. (2008). Use of a multiplexed oscillator in a miniaturized electronic nose based on a multichannel quartz crystal microbalance. *Sensors and actuators B: Chemical, 131*(1), 159-166.

Zane L, Bargelloni L, Patarnello T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology, 11*(1), 1-16.

Zareef M, Chen Q, Ouyang Q, Kutsanedzie FYH, Hassan MM, Viswadevarayalu A, et al. (2018). Prediction of amino acids, caffeine, theaflavins and water extract in black tea using FT-NIR spectroscopy coupled chemometrics algorithms. *Analytical Methods, 10*(25), 3023-3031. doi:10.1039/C8AY00731D

Zeeb DJ, Nelson BC, Albert K, Dalluge JJ. (2000). Separation and identification of twelve catechins in tea using liquid chromatography/atmospheric pressure chemical ionization-mass spectrometry. *Analytical Chemistry, 72*(20), 5020-5026.

Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, et al. (2017a). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in plant science, 8*, 1916.

Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, et al. (2017b). Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Frontiers in Plant Science, 8*, 1916-1916. doi:10.3389/fpls.2017.01916

Zhang J, Mason AS, Wu J, Liu S, Zhang X, Luo T, et al. (2015). Identification of putative candidate genes for water stress tolerance in canola (Brassica napus). *Frontiers in Plant Science, 6*, 1058.

Zhang L, Tai Y, Wang Y, Meng Q, Yang Y, Zhang S, et al. (2017c). The proposed biosynthesis of procyanidins by the comparative chemical analysis of five Camellia species using LC-MS. *Sci Rep, 7*, 46131. doi:10.1038/srep46131

Zhang X, Mobley N, Zhang J, Zheng X, Lu L, Ragin O, et al. (2010). Analysis of agricultural residues on tea using d-SPE sample preparation with GC-NCI-MS and UHPLC-MS/MS. *Journal of agricultural and food chemistry, 58*(22), 11553-11560.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. (2007a). An Arabidopsis example of association mapping in structured samples. *PLoS genetics, 3*(1), e4.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. (2007b). An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS genetics, 3*(1), e4. doi:10.1371/journal.pgen.0030004

Zheng X-Q, Li Q-S, Xiang L-P, Liang Y-R. (2016a). Recent advances in volatiles of teas. *Molecules, 21*(3), 338.

Zheng X-Q, Li Q-S, Xiang L-P, Liang Y-R. (2016b). Recent Advances in Volatiles of Teas. *Molecules, 21*(3). doi:10.3390/molecules21030338

Zhi R, Zhao L, Shi B, Jin Y. (2014). New dimensionality reduction model (manifold learning) coupled with electronic tongue for green tea grade identification. *European Food Research and Technology, 239*(1), 157-167.

Zhi R, Zhao L, Zhang D. (2017). A framework for the multi-level fusion of electronic nose and electronic tongue for tea quality assessment. *Sensors, 17*(5), 1007.

Zhong S, Dekkers JCM, Fernando RL, Jannink J-L. (2009a). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics, 182*(1), 355-364.

Zhong S, Dekkers JCM, Fernando RL, Jannink JL. (2009b). Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics, 182*(1), 355-364. doi:10.1534/genetics.108.098277

Zhou Q, Zhou C, Zheng W, Mason AS, Fan S, Wu C, et al. (2017). Genome-wide SNP markers based on SLAF-seq uncover breeding traces in rapeseed (Brassica napus L.). *Frontiers in Plant Science, 8*, 648.

Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK, Smith NJ. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant breeding, 122*(1), 40-46.

Zhu C, Gore M, Buckler ES, Yu J. (2008). Status and prospects of association mapping in plants. *The plant genome, 1*(1), 5-20.

Zhu H, Ye Y, He H, Dong C. (2017). Evaluation of green tea sensory quality via process characteristics and image information. *Food and Bioproducts Processing, 102*, 116-122. doi:10.1016/j.fbp.2016.12.004

Zhu X-M, Shao X-Y, Pei Y-H, Guo X-M, Li J, Song X-Y, et al. (2018). Genetic Diversity and Genome-Wide Association Study of Major Ear Quantitative Traits Using High-Density SNPs in Maize. *Frontiers in Plant Science, 9*(966). doi:10.3389/fpls.2018.00966

Zhuang X, Wang L, Chen Q, Wu X, Fang J. (2017). Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools. *Science China Technological Sciences, 60*(1), 84-90. doi:10.1007/s11431-016-0464-0

Zoecklein BW, Devarajan YS, Mallikarjunan K, Gardner DM. (2011). Monitoring effects of ethanol spray on Cabernet franc and Merlot grapes and wine volatiles using electronic nose systems. *American journal of enology and viticulture*, ajev-2011.

Zou G, Xiao Y, Wang M, Zhang H. (2018). Detection of bitterness and astringency of green tea with different taste by electronic nose and tongue. *PLoS One, 13*(12), e0206517.

Zuo Y, Chen H, Deng Y. (2002). Simultaneous determination of catechins, caffeine and gallic acids in green, Oolong, black and pu-erh teas using HPLC with a photodiode array detector. *Talanta, 57*(2), 307-316.

Table S7. 1. Mean biochemical traits (mg per gram tissue) across Jamji and Kericho. The values colored green indicates not significantly (p<0.05) different from each other for each trait.

	clone	TB	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
1	CLONE97	3.52	37.49	24.39	4.85	25.00	24.71	112.61	20.67	72.75	11.76	57.40	21.76
2	CLONE101	4.98	36.06	20.71	5.29	30.82	25.54	81.31	46.86	91.77	8.57	39.90	13.86
3	Clone69	3.98	39.88	22.38	5.73	31.57	28.11	89.62	40.77	90.22	5.84	34.59	22.11
4	Clone70	2.57	35.48	21.73	5.82	28.41	33.01	97.77	30.59	68.34	8.76	50.85	14.18
5	Clone37	3.55	39.17	18.96	4.85	34.67	22.20	65.05	51.35	106.51	5.46	31.81	13.29
6	Clone89	4.69	38.40	15.38	4.13	32.53	20.91	59.85	53.05	101.46	7.52	39.38	20.44
7	Clone79	4.51	36.72	16.55	4.58	31.11	21.74	62.95	53.73	97.87	4.47	40.89	21.25
8	Clone61	6.84	41.59	17.96	3.54	29.17	16.15	71.80	25.78	93.42	12.78	63.28	14.41
9	Clone25	2.10	29.67	14.68	5.06	28.76	27.02	64.14	69.30	77.97	5.00	39.44	15.34
10	Clone55	5.78	44.18	16.39	4.21	34.95	17.99	60.22	40.04	120.82	4.25	33.21	15.07
11	Clone57	5.82	41.38	14.43	4.66	30.38	23.92	64.42	56.30	92.37	4.49	38.60	19.66
12	Clone68	6.24	40.35	18.25	4.07	34.47	18.59	71.77	36.16	102.91	4.37	41.47	21.34
13	Clone87	4.81	42.47	15.69	4.47	29.87	23.41	70.88	50.06	85.22	3.75	44.39	20.38
14	Clone81	5.52	40.40	15.32	3.36	24.99	18.29	72.88	39.73	75.46	12.82	63.10	29.82
15	Clone63	6.17	36.94	14.79	4.51	32.59	19.73	57.12	50.69	106.27	4.89	35.86	18.46
16	Clone53	3.02	33.57	21.40	4.98	28.38	24.27	93.17	26.02	74.71	8.54	48.94	21.18
17	CLONE95	6.70	41.15	18.11	4.52	31.65	20.80	61.03	46.69	101.11	6.27	35.82	21.54
18	Clone62	4.51	42.73	15.43	4.55	33.35	22.13	54.82	52.07	108.96	3.47	28.01	17.18
19	CLONE102	6.49	42.48	15.24	4.76	29.91	22.13	58.27	49.12	103.06	4.53	35.67	15.60
20	Clone56	5.85	37.17	18.91	3.66	28.67	16.74	85.83	17.13	91.86	10.89	51.40	19.81
21	Clone71	3.96	43.40	15.37	4.32	31.14	20.51	55.24	51.08	107.86	3.31	33.10	16.70
22	Clone60	5.40	37.17	15.78	4.30	30.81	20.08	55.63	49.92	102.68	5.75	35.95	18.66
23	Clone46	3.93	42.06	14.60	4.27	29.76	22.20	62.02	51.28	93.39	5.72	35.68	18.87
24	Clone9	4.21	37.69	16.80	3.83	25.67	20.21	75.69	37.84	77.76	9.67	53.11	20.46
25	Clone3	3.47	35.98	15.26	3.72	32.43	18.26	58.25	50.80	103.52	5.55	30.50	14.31
26	Clone67	4.69	42.11	13.75	3.98	29.32	17.18	50.26	49.28	106.93	7.19	38.57	24.64
27	Clone18	3.72	38.69	16.18	4.39	31.37	23.06	57.63	58.73	93.61	5.93	28.01	20.76
28	Clone21	3.16	36.87	15.66	3.58	25.13	19.20	73.91	35.65	77.41	11.04	55.87	19.73
29	Clone29	3.29	35.90	15.09	4.91	32.52	22.36	50.95	62.96	99.84	2.93	26.27	14.97
30	Clone91	3.20	34.28	18.59	4.30	31.24	21.92	69.85	39.70	93.64	4.68	35.47	22.72
31	Clone31	5.75	33.49	18.41	3.47	26.62	17.58	75.49	33.20	74.42	11.31	56.46	20.45
32	Clone49	2.72	44.02	14.85	4.30	31.62	21.59	60.95	47.57	99.29	7.26	26.74	18.57
33	Clone92	6.21	43.00	18.99	5.33	32.97	26.90	65.73	47.57	91.24	4.62	25.49	13.89
34	Clone64	5.58	42.79	15.03	3.99	30.87	18.16	59.64	40.53	96.60	6.45	41.52	14.03
35	CLONE100	4.13	37.52	15.17	3.89	30.61	17.42	53.83	47.72	101.45	3.41	39.23	20.40
36	Clone66	3.77	36.97	14.40	4.65	32.72	21.04	48.59	57.72	99.30	4.68	28.41	10.90
37	Clone51	4.38	40.08	17.30	4.51	32.37	21.69	61.43	41.18	91.68	5.27	37.01	18.89
38	Clone17	4.05	37.46	16.22	3.47	26.12	18.13	68.94	35.53	76.51	9.22	56.09	22.37
39	Clone65	2.87	33.69	13.93	3.69	32.10	19.46	46.39	58.87	100.58	4.89	26.94	10.70
40	Clone48	5.70	46.66	17.55	4.03	33.47	19.62	53.86	42.32	106.72	3.79	29.34	14.83
41	Clone82	3.42	36.08	17.52	4.48	31.29	22.59	63.96	48.06	89.71	4.47	28.71	21.00
42	Clone72	4.12	28.36	16.50	5.14	31.02	26.16	55.79	64.42	82.80	4.03	23.97	12.09

43	Clone16	3.15	43.07	16.70	3.96	28.05	19.25	73.60	29.22	84.27	7.39	46.35	21.49
44	Clone94	2.88	36.70	15.52	4.52	26.57	23.35	75.72	32.71	69.43	8.64	51.45	13.94
45	Clone30	4.53	38.41	13.21	3.45	28.96	16.44	49.71	51.18	94.93	5.43	39.77	16.17
46	Clone33	4.11	38.00	13.15	4.23	29.40	21.90	54.58	50.61	90.09	4.65	34.47	16.56
47	CLONE103	3.14	33.63	13.33	4.18	30.92	20.20	47.10	58.94	89.81	4.34	33.67	17.47
48	Clone5	3.69	38.08	16.85	3.50	24.33	16.69	77.42	25.16	73.80	7.42	59.73	17.01
49	Clone86	6.26	42.63	17.16	4.98	28.95	25.23	64.71	45.55	81.23	6.56	31.62	26.21
50	Clone13	3.94	37.20	12.90	4.03	30.17	16.79	44.71	59.17	92.73	5.30	34.57	17.31
51	Clone34	3.66	37.99	18.14	4.39	25.09	19.97	68.56	35.12	73.25	7.80	53.59	20.43
52	Clone2	3.47	34.75	18.83	3.76	23.73	17.78	84.34	22.72	66.08	10.45	58.17	14.46
53	Clone90	4.12	41.17	14.16	3.91	31.32	18.96	49.51	49.52	97.52	5.46	30.83	16.56
54	Clone85	4.32	33.89	14.12	3.14	31.30	16.05	40.69	60.15	101.46	3.71	29.61	19.18
55	CLONE98	4.79	39.95	16.59	3.37	25.62	17.44	74.42	26.95	79.13	7.76	51.28	19.13
56	Clone14	2.88	36.22	13.03	3.82	28.77	18.89	49.54	57.70	92.98	3.91	30.05	16.95
57	Clone44	4.10	35.74	15.98	3.36	27.76	19.29	63.87	38.74	82.88	6.34	41.28	25.48
58	Clone8	4.77	38.99	17.07	3.34	28.53	15.86	64.85	26.79	91.82	7.64	43.94	18.78
59	Clone54	6.36	39.41	16.23	3.79	29.63	18.68	60.51	36.67	93.61	6.36	33.63	18.53
60	Clone36	3.06	34.39	16.05	4.03	26.01	19.83	70.75	34.00	76.65	7.20	44.22	18.97
61	Clone19	3.17	38.32	13.57	3.89	28.47	18.58	54.81	48.60	86.22	5.88	36.07	18.42
62	Clone40	4.66	39.64	19.03	3.61	28.54	18.21	71.95	27.30	83.04	7.29	41.33	17.92
63	Clone38	4.40	37.18	14.06	3.58	29.12	16.69	53.26	43.13	94.60	6.10	34.77	21.42
64	Clone7	4.22	36.08	17.07	2.99	26.03	14.76	67.07	26.58	84.42	9.19	49.55	19.27
65	Clone83	4.86	33.84	14.15	3.59	27.85	18.89	54.69	50.16	86.10	6.15	33.35	23.57
66	Clone32	3.62	42.31	15.99	4.08	29.92	18.43	49.31	49.21	94.21	5.45	30.04	15.56
67	Clone42	6.33	40.71	12.85	3.88	27.18	19.17	46.59	55.58	89.45	6.18	32.41	20.11
68	Clone12	3.88	37.88	17.81	4.55	30.47	21.54	57.59	42.67	94.86	3.83	25.46	20.36
69	Clone50	6.10	37.46	17.10	4.36	27.93	21.55	57.68	47.18	87.40	4.00	30.43	16.82
70	Clone78	4.94	37.06	14.21	3.69	26.73	20.59	54.83	50.90	81.68	6.71	34.64	25.80
71	Clone22	4.46	41.78	15.96	2.94	27.46	13.93	59.23	27.37	91.94	8.37	47.36	24.72
72	Clone75	4.25	41.46	12.48	3.45	27.14	17.21	50.90	48.38	79.68	7.31	44.99	26.71
73	Clone93	4.50	39.67	15.45	3.92	27.46	20.75	62.03	46.49	72.85	7.01	38.27	18.43
74	Clone4	5.60	42.46	16.83	3.67	27.18	16.72	62.90	28.64	82.14	9.20	47.47	17.57
75	Clone28	5.04	43.06	16.80	3.80	26.61	19.16	75.24	23.39	73.45	7.95	48.09	23.33
76	Clone1	5.67	38.04	16.77	4.02	29.95	17.71	53.38	43.84	91.29	4.77	32.70	18.64
77	Clone76	3.82	33.75	16.59	4.64	30.98	24.37	57.48	51.11	74.90	3.96	30.24	15.65
78	Clone6	3.08	34.49	13.24	3.90	26.88	19.64	52.05	51.36	84.08	5.51	32.43	16.02
79	Clone35	4.84	40.24	15.45	3.67	28.57	17.33	62.86	33.91	79.16	6.16	43.61	18.02
80	Clone58	4.52	33.22	16.27	4.59	27.59	23.32	61.65	41.78	77.49	6.31	31.98	20.00
81	Clone27	3.44	39.43	12.63	3.88	29.74	18.32	43.98	54.63	97.16	3.89	22.36	19.39
82	CLONE99	3.63	34.34	13.24	4.24	27.68	20.78	48.21	52.26	86.77	4.35	29.92	17.13
83	Clone73	4.40	37.00	14.69	4.46	28.60	20.68	48.57	51.51	89.03	3.93	27.51	21.72
84	Clone45	2.96	33.35	11.64	3.84	22.87	19.46	59.12	39.03	59.89	9.11	60.17	16.96
85	Clone23	2.20	34.67	15.48	3.69	22.83	18.85	72.18	24.92	67.55	9.64	51.13	20.81
86	Clone20	2.60	32.88	12.99	5.66	28.14	27.74	54.13	52.13	67.11	5.06	31.06	19.95
87	Clone52	3.71	41.51	13.22	4.05	28.42	18.11	46.66	51.11	84.13	4.78	31.71	16.77
88	Clone84	5.31	35.40	14.22	4.06	31.70	18.26	45.02	48.00	94.25	3.47	24.02	29.10
89	Clone88	4.45	37.98	12.43	4.32	28.29	22.58	44.38	65.02	70.83	4.03	28.93	17.89
90	Clone39	2.53	35.93	17.56	3.12	25.24	14.37	68.35	18.53	83.96	6.84	46.58	12.23
91	Clone10	2.68	27.92	14.87	3.79	28.13	18.32	47.74	48.00	90.35	3.24	25.91	16.77
92	Clone26	2.44	30.84	15.47	3.80	27.44	17.91	47.37	47.56	86.40	4.16	30.20	15.05
93	Clone43	3.58	36.58	16.83	3.50	25.61	16.27	60.75	29.58	74.48	8.34	45.06	22.49
94	Clone59	6.13	39.54	17.99	3.63	27.92	17.99	59.50	30.46	87.39	4.57	31.45	19.25
95	Clone77	3.86	35.24	14.02	3.29	27.30	15.54	47.83	42.63	86.61	5.91	31.65	17.95
96	Clone24	3.48	35.61	11.47	4.72	25.01	24.12	48.65	57.54	68.92	4.83	26.99	17.30

97	Clone47	5.62	40.19	15.46	3.78	28.07	18.47	49.03	40.65	87.92	3.37	27.80	22.32
98	Clone41	3.97	39.36	18.03	3.38	27.20	15.64	63.69	23.76	77.06	6.15	40.76	25.69
99	Clone80	6.51	40.77	15.79	3.55	28.66	17.02	45.40	38.31	94.13	4.29	22.70	27.11
100	Clone74	3.83	34.15	15.03	4.35	26.75	20.95	45.47	51.77	72.73	4.12	26.51	26.23
101	Clone11	2.41	35.11	12.23	3.34	25.28	16.08	40.51	49.83	80.27	4.10	27.12	28.85
102	Clone15	3.44	37.61	12.88	3.04	22.51	13.65	53.91	26.95	66.99	7.10	44.47	21.71
103	CLONE96	2.85	30.37	11.67	4.27	24.21	22.04	46.74	41.59	66.51	4.88	26.85	13.70

Table S7. 2. Mean biochemical values (mg per gram tissue) of the NMR spectroscopy for the technical replicates of all the traits.

Biological Replicate	Technical replicate	Clone	TB	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
1	1	CLONE97	0.26	27.59	13.31	2.24	12.77	23.30	99.80	9.09	54.69	17.65	51.45	14.39
1	1	Clone22	0.93	33.25	17.89	2.27	10.86	10.14	49.56	18.93	70.96	15.12	37.87	13.81
1	1	CLONE99	1.20	32.77	10.89	2.88	20.66	13.06	42.38	36.64	71.87	3.90	21.35	14.12
1	1	Clone65	0.14	28.44	14.41	2.45	20.39	12.50	40.26	44.19	76.58	5.71	17.45	8.18
1	1	Clone21	0.72	28.48	12.37	1.32	18.80	12.14	63.97	26.58	56.99	5.95	43.54	16.63
1	1	Clone11	0.60	33.07	14.12	3.26	19.67	10.55	35.09	43.75	67.96	5.68	16.44	20.33
1	1	Clone64	2.32	42.64	13.17	4.82	17.67	13.96	50.90	25.19	84.22	5.18	28.66	10.12
1	1	Clone37	0.88	38.38	19.46	5.28	23.04	16.20	54.07	34.76	90.13	4.03	18.86	13.11
1	1	CLONE96	1.05	29.69	8.83	2.58	26.24	14.53	45.07	34.14	54.26	3.18	15.37	9.02
1	1	Clone93	1.18	35.45	13.37	2.34	21.81	17.18	68.87	32.62	56.08	21.42	40.32	8.71
1	1	Clone26	0.81	30.84	19.33	3.19	16.82	12.93	41.78	28.79	67.29	3.28	16.49	13.68
1	1	Clone58	0.79	31.60	11.16	2.64	24.37	17.57	58.52	40.14	73.20	5.00	24.26	12.82
1	1	Clone78	2.14	36.63	12.83	1.65	23.61	15.83	45.62	44.32	67.03	3.18	18.17	16.12
1	1	Clone89	2.50	36.20	10.39	2.42	20.95	15.36	51.69	43.48	80.38	12.02	28.37	16.48
1	1	Clone42	2.00	36.79	12.81	2.89	19.15	13.20	49.80	41.15	76.05	12.82	33.04	10.14
1	1	Clone88	2.77	33.65	7.71	1.63	25.91	11.30	38.45	47.82	56.63	3.85	18.93	16.05
1	1	Clone19	0.78	32.11	10.37	2.65	19.13	12.11	42.34	39.87	67.60	4.69	22.82	12.30
1	1	Clone73	0.42	33.93	10.99	2.89	18.94	11.90	41.59	36.12	74.50	2.81	17.88	16.39
1	1	#N/A	1.55	33.78	14.79	2.44	16.66	11.98	53.65	26.68	65.85	12.17	33.39	11.63
1	1	CLONE98	0.55	29.52	13.38	0.90	11.11	11.31	60.34	16.21	59.65	7.75	37.39	13.13
1	1	CLONE101	0.89	31.92	16.29	3.38	21.72	18.45	62.63	32.41	71.40	10.87	24.73	8.00
1	1	Clone8	1.76	38.49	19.08	3.14	11.81	12.56	58.40	14.12	78.60	4.11	36.08	18.48
1	1	Clone28	1.75	40.62	14.49	3.39	19.69	16.96	71.83	17.47	57.03	6.20	37.05	18.14
1	1	Clone77	1.21	30.29	12.97	2.04	12.86	9.40	35.80	30.51	64.72	6.78	22.41	10.79
1	1	Clone18	0.53	31.96	10.14	1.97	21.29	13.50	44.42	43.54	71.47	7.03	17.38	15.53
1	1	Clone1	2.19	35.87	16.90	2.80	16.41	13.93	46.02	30.35	75.55	3.33	23.55	12.22
1	1	Clone60	0.00	33.65	11.03	2.35	16.66	11.27	39.29	40.94	74.55	3.11	23.82	18.47
1	1	CLONE95	2.70	39.31	16.96	3.54	17.72	15.06	50.96	32.31	90.17	12.11	27.08	13.43
1	1	CLONE104	1.26	30.65	10.49	2.73	11.59	8.08	26.17	34.97	58.80	3.65	14.92	9.19
1	1	CLONE105	2.13	26.09	12.92	1.90	21.31	10.71	42.00	60.13	76.51	6.86	28.05	14.33
1	1	TRFK 301/6	7.40	16.87	18.42	7.25	26.44	21.33	59.26	18.40	84.21	1.73	0.91	7.71
1	2	CLONE97	0.33	28.50	13.94	2.28	14.13	23.57	102.62	10.05	54.65	7.33	47.21	15.36
1	2	Clone22	1.14	38.57	19.50	2.59	12.59	12.44	55.05	20.70	75.82	12.57	37.72	14.59
1	2	CLONE99	1.14	33.33	10.82	2.62	20.17	13.83	43.25	36.29	72.96	2.68	20.73	15.01
1	2	Clone65	0.35	27.06	13.87	1.88	18.58	12.13	40.60	40.90	73.24	2.56	17.59	8.17
1	2	Clone21	0.97	29.97	12.86	1.51	19.32	13.95	65.47	27.29	58.87	14.82	44.78	17.79
1	2	Clone11	0.65	34.50	14.62	2.75	19.62	11.56	37.98	45.49	72.39	4.07	17.37	20.99
1	2	Clone64	2.31	41.76	13.00	4.47	16.85	13.71	51.45	25.18	85.45	4.29	28.08	10.04
1	2	Clone37	1.08	35.91	19.31	3.98	20.15	14.73	51.63	34.38	83.84	4.67	17.69	12.32
1	2	CLONE96	1.24	30.25	9.17	3.12	28.20	15.21	45.05	34.91	55.78	3.73	14.37	9.08
1	2	Clone93	1.02	33.88	13.11	1.92	20.93	16.96	71.45	33.28	59.31	24.89	47.08	9.15
1	2	Clone26	0.53	29.87	18.79	3.17	16.09	11.97	42.18	28.05	67.28	3.30	18.12	13.68
1	2	Clone58	0.65	32.59	11.93	3.64	27.17	17.93	59.69	41.22	75.17	5.71	25.81	13.13

1	2	Clone78	1.86	34.06	12.00	1.25	22.10	14.29	44.74	44.60	66.31	10.03	22.02	15.93
1	2	Clone89	2.75	37.10	11.10	2.61	21.96	16.82	53.10	45.42	82.39	8.75	27.34	16.18
1	2	Clone42	1.65	35.52	12.51	2.51	17.93	13.19	50.76	40.82	77.89	7.61	33.12	9.96
1	2	Clone88	1.39	34.22	7.33	1.42	25.85	10.99	39.22	48.45	60.79	2.89	19.41	15.82
1	2	Clone19	0.74	34.21	11.16	2.77	20.42	12.32	47.39	43.19	74.81	9.25	25.85	13.15
1	2	Clone73	0.89	32.21	14.17	2.34	15.64	10.73	51.89	25.45	63.67	7.72	33.52	11.44
1	2	Clone36	1.58	33.88	14.92	2.68	17.08	11.90	53.47	26.46	66.47	12.22	33.58	11.52
1	2	CLONE98	1.02	33.16	15.12	1.96	15.35	14.54	63.07	19.75	61.23	19.35	34.46	13.13
1	2	CLONE101	1.53	32.52	16.91	3.65	22.59	19.12	63.86	31.74	72.17	3.25	23.84	7.86
1	2	Clone8	1.53	38.27	18.66	3.25	11.88	12.43	58.35	14.30	77.52	4.12	37.47	18.72
1	2	Clone28	1.83	40.18	14.78	3.11	19.46	17.28	72.22	18.08	57.45	7.54	37.99	17.87
1	2	Clone77	1.49	32.69	13.85	2.10	13.89	10.40	37.93	31.88	67.81	4.84	21.51	11.04
1	2	Clone18	0.12	31.91	10.27	1.73	20.74	13.97	45.95	42.96	71.05	6.05	18.01	15.13
1	2	Clone1	1.33	35.05	16.65	2.39	15.40	13.36	44.26	30.33	74.88	9.46	24.30	12.09
1	2	Clone60	0.21	34.19	11.09	2.48	16.94	11.01	39.09	40.78	73.51	5.11	21.83	18.43
1	2	CLONE95	2.51	40.71	17.65	4.13	19.45	15.16	51.04	32.05	90.19	4.13	24.24	13.54
1	2	CLONE104	1.13	33.04	11.62	3.15	13.40	9.30	27.58	36.11	63.86	3.60	14.65	9.91
1	2	CLONE105	2.03	26.32	13.02	2.52	22.46	10.65	40.77	59.64	73.04	5.48	27.68	14.05
1	2	CLONE106	7.35	17.20	21.77	8.96	32.14	24.89	71.40	22.59	102.11	1.99	1.05	8.75
1	3	CLONE97	0.70	31.37	15.72	2.71	17.01	27.35	103.61	11.62	57.61	23.56	45.68	15.15
1	3	Clone22	1.28	38.06	19.46	2.60	12.39	12.38	53.36	20.23	74.29	6.38	37.29	14.38
1	3	CLONE99	1.30	32.48	10.84	3.13	21.08	12.88	41.60	35.87	70.58	7.83	20.57	14.39
1	3	Clone65	0.31	27.93	14.12	1.95	19.13	12.35	40.58	42.88	74.60	6.09	17.03	8.19
1	3	Clone21	0.89	29.76	12.69	1.49	19.50	13.55	63.42	27.50	57.21	24.93	43.78	17.97
1	3	Clone11	0.59	31.50	13.19	2.38	17.42	10.48	35.33	41.21	68.38	5.00	18.72	20.41
1	3	Clone64	2.05	39.48	11.89	4.51	16.16	11.53	48.56	24.86	83.93	6.55	27.70	9.69
1	3	Clone37	1.36	39.83	20.79	4.82	22.50	17.12	54.46	35.88	90.66	6.79	18.63	12.81
1	3	CLONE96	0.94	30.05	9.03	2.89	27.96	14.95	46.48	34.39	57.79	4.56	16.48	9.29
1	3	Clone93	0.98	32.78	12.58	1.74	19.15	16.58	69.00	31.66	57.40	8.13	45.49	9.16
1	3	Clone26	1.00	32.52	20.13	3.41	17.54	15.00	43.94	29.15	71.16	3.02	16.69	14.69
1	3	Clone58	0.18	31.55	11.92	3.17	26.04	17.22	57.49	41.15	71.42	5.25	24.01	12.85
1	3	Clone78	1.77	33.09	11.44	1.30	21.33	13.16	43.20	42.75	63.57	9.72	21.29	15.95
1	3	Clone89	2.52	34.49	9.95	2.40	20.68	14.25	51.54	43.43	80.62	15.77	31.25	16.62
1	3	Clone42	1.86	37.62	13.43	3.07	19.41	13.96	50.19	42.10	76.44	10.56	29.69	9.95
1	3	Clone88	2.81	35.26	7.95	1.95	27.56	11.59	38.99	50.26	59.20	3.68	19.39	16.22
1	3	Clone19	1.16	37.71	12.43	4.07	23.69	14.42	45.82	43.98	71.35	4.15	17.57	12.15
1	3	Clone73	0.20	34.33	11.39	3.20	20.12	13.66	42.37	35.18	76.19	2.56	18.21	16.37
1	3	Clone36	1.51	33.67	14.66	2.20	16.24	12.07	53.82	26.90	65.23	12.12	33.20	11.75
1	3	CLONE98	0.99	32.88	15.03	1.88	14.98	14.25	63.55	19.63	61.99	17.44	35.30	13.36
1	3	CLONE101	1.47	33.55	16.82	4.14	23.65	18.93	63.74	32.03	72.39	3.45	24.55	8.12
1	3	Clone8	1.78	38.11	19.75	3.39	12.74	13.06	58.40	16.12	79.22	13.81	35.43	18.61
1	3	Clone28	1.89	41.17	15.34	3.55	20.93	17.25	71.11	18.46	56.71	6.87	37.49	17.92
1	3	Clone77	1.41	32.83	14.21	2.14	14.40	10.66	37.77	32.32	67.56	3.03	20.95	10.79
1	3	Clone18	0.24	32.86	10.54	1.91	21.11	14.46	46.52	44.30	73.65	7.71	18.83	15.94
1	3	Clone1	1.20	37.60	17.78	3.17	16.96	12.68	45.06	31.78	76.83	3.88	20.49	11.98
1	3	Clone60	0.43	36.15	12.08	2.63	17.88	11.22	37.97	43.88	70.35	6.86	16.38	17.53
1	3	CLONE95	2.66	41.64	17.83	4.07	19.30	16.04	51.88	32.37	90.12	3.36	25.52	13.71
1	3	CLONE104	0.74	31.85	10.73	2.97	12.01	7.90	27.52	33.83	61.92	6.20	16.91	9.85
1	3	CLONE105	1.80	26.95	13.20	2.35	22.08	11.12	40.60	60.41	68.83	11.41	22.43	13.48
1	3	CLONE106	6.84	16.34	18.84	7.50	27.03	21.52	63.74	18.51	90.80	1.80	0.95	8.09

Table S7. 3. Mean squares of ANOVA for biochemical compounds across the two sites.

	TB	Caffeine	Theogallin	GA	EC	GCG	ECG	EGC	EGCG	Catechin	GC	Theanine
Genotype	8.21***	77.82***	33.02***	2.1648***	44.59**	66.51***	977.99***	809.26***	844.8***	30.2***	616.04***	94.18***
Environment	160.77***	433.98***	406.83***	28.65***	341.05***	18.3ns	1623.19***	558.94**	3936.8***	0.42*	13.25 ns	666.04***
G x E	2.1 ns	13.93 ns	5.54 ns	0.69 ns	16.83 ns	14.44 ns	86.29 ns	63.27 ns	136.7 ns	10.5 ns	88.41***	38.09 ns

Error	2.74	14.48	8.56	0.86	31.87	13.06	68.83	67.26	163.8	12.09	48.67	32.59
CV (%)	38.7	10.1	18.5	22.8	19.6	18.1	13.7	18.9	14.7	56.5	18.4	29.9
LSD (p<0.05)	2.66	6.107	4.7	1.49	9.06	5.8	13.316	13.16	20.54	5.58	11.20	9.163

* Significant at P < 0.05.
** Significant at P < 0.01.
*** Significant at P < 0.001.
ns, not significant, GCG, gallocatechin gallate, ECG, epicatechin gallate, EGC, epigallocatechin, EGCG. Epigallocatechin gallate, GC, gallocatechin.

Table S7. 4. Significant SNP markers for all the biochemical traits.

	Trait	Marker	Gene ID	p-Value	R2
1	Caffeine	TP394	TEA028063.1	8.99E-05	0.20
2	Catechin	TP32223	TEA028480.1	2.22E-04	0.17
3	Catechin	TP32224	TEA017847.1	2.63E-04	0.17
4	ECG	TP10709	TEA012849.1	3.08E-05	0.19
5	ECG	TP12308	TEA015664.1	1.56E-04	0.23
6	ECG	TP13032	TEA029936.1	1.26E-05	0.37
7	ECG	TP14758	TEA021748.1	8.97E-05	0.24
8	ECG	TP15975	TEA011547.1	2.77E-04	0.13
9	ECG	TP19784	TEA016025.1	2.71E-06	0.23
10	ECG	TP27846	TEA031023.1	3.12E-04	0.17
11	ECG	TP28250	TEA016188.1	2.21E-05	0.23
12	ECG	TP32223	TEA028480.1	2.65E-04	0.17
13	ECG	TP32224	TEA017847.1	3.13E-04	0.17
14	ECG	TP36831	TEA032612.1	1.32E-06	0.30
15	ECG	TP39815	TEA032995.1	2.29E-05	0.34
16	ECG	TP45007	TEA014458.1	1.13E-05	0.37
17	ECG	TP47049	TEA010521.1	3.40E-05	0.28
18	ECG	TP495	TEA026420.1	2.50E-04	0.17
19	ECG	TP49950	TEA002229.1	2.12E-07	0.29
20	ECG	TP59980	TEA022841.1	2.24E-05	0.23
21	ECG	TP59981	TEA033170.1	2.24E-05	0.23
22	ECG	TP62088	TEA001150.1	1.02E-04	0.21
23	ECG	TP67837	TEA018961.1	6.26E-07	0.34
24	ECG	TP71141	TEA012194.1	6.57E-06	0.42
25	ECG	TP72124	TEA030042.1	1.71E-05	0.26
26	ECG	TP72125	TEA018384.1	9.98E-06	0.29
27	ECG	TP75665	TEA026342.1	2.24E-05	0.23
28	ECG	TP75737	TEA029957.1	2.14E-04	0.21
29	ECG	TP77023	TEA019659.1	2.86E-04	0.20
30	ECG	TP81402	TEA002759.1	9.25E-05	0.24
31	ECG	TP81945	TEA008015.1	1.27E-04	0.22
32	ECG	TP8361	TEA021225.1	2.82E-04	0.32
33	EGC	TP71786	TEA022074.1	4.45E-05	0.20
34	Gallic Acid	TP16595	TEA008540.1	8.43E-05	0.20
35	GC	TP71786	TEA022074.1	2.87E-04	0.17
36	GCG	TP15601	TEA020725.1	2.33E-04	0.21

37	GCG	TP16595	TEA008540.1	8.65E-06	0.26
38	GCG	TP27072	TEA003563.1	7.53E-05	0.36
39	GCG	TP28250	TEA016188.1	7.97E-05	0.20
40	GCG	TP45007	TEA014458.1	2.11E-04	0.35
41	GCG	TP47049	TEA010521.1	1.32E-04	0.23
42	GCG	TP49950	TEA002229.1	2.00E-04	0.15
43	GCG	TP53892	TEA009955.1	5.20E-05	0.24
44	GCG	TP63459	TEA007457.1	3.36E-05	0.48
45	GCG	TP64501	TEA012672.1	1.92E-04	0.19
46	GCG	TP67837	TEA018961.1	1.18E-04	0.21
47	GCG	TP69675	TEA011118.1	1.42E-04	0.22
48	GCG	TP71141	TEA012194.1	1.78E-04	0.30
49	GCG	TP71587	TEA006550.1	2.12E-04	0.19
50	GCG	TP81516	TEA016098.1	1.61E-04	0.19
51	GCG	TP9296	TEA031431.1	9.88E-05	0.18
52	GCG	TP957	TEA003188.1	1.51E-04	0.22
53	Theobromine	TP394	TEA028063.1	7.61E-06	0.26
54	Theobromine	TP8427	TEA030743.1	2.57E-04	0.14
55	Theogallin	TP10709	TEA012849.1	1.35E-04	0.16
56	Theogallin	TP19784	TEA016025.1	1.10E-05	0.20
57	Theogallin	TP28250	TEA016188.1	2.62E-04	0.17
58	Theogallin	TP39815	TEA032995.1	2.49E-05	0.37
59	Theogallin	TP495	TEA026420.1	1.79E-04	0.18
60	Theogallin	TP49950	TEA002229.1	1.18E-04	0.15
61	Theogallin	TP67837	TEA018961.1	6.50E-06	0.29
62	Theogallin	TP71141	TEA012194.1	7.33E-05	0.28
63	Theogallin	TP72125	TEA018384.1	3.06E-04	0.21
64	Theogallin	TP78835	TEA018622.1	2.23E-04	0.18

Table S7. 5. KEGG ontology pathways.

	Pathway	Pathway ID	No. enzyme in pathway	Enzyme
1	Glycolysis / Gluconeogenesis	map00010	2	ec:4.1.1.32 - carboxykinase
2	Glyoxylate and dicarboxylate metabolism	map00630	2	ec:4.1.3.1 - lyase, ec:6.2.1.1 - ligase
3	Glutathione metabolism	map00480	4	ec:1.11.1.15 - thioredoxin peroxidase, ec:1.17.4.1 - reductase, ec:1.1.1.49 - dehydrogenase (NADP+), ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating)
4	Tryptophan metabolism	map00380	1	ec:3.5.1.4 - acylamidase
5	Pyruvate metabolism	map00620	4	ec:4.1.1.32 - carboxykinase
6	Caffeine metabolism	map00232	2	ec:1.17.3.2 - oxidase, ec:2.3.1.5 - N-acetyltransferase
7	Biosynthesis of antibiotics	map01130	13	ec:2.6.1.42 - transaminase, ec:2.3.1.1 - N-acetyltransferase, ec:1.3.5.1 - dehydrogenase, ec:4.1.1.32 - carboxykinase (GTP), ec:4.2.1.9 - dehydratase, ec:6.4.1.2 - carboxylase, ec:1.1.1.49 - dehydrogenase (NADP+), ec:6.3.5.3 - synthase, ec:2.3.1.35 - N-acetyltransferase, ec:2.7.2.4 -
8	Monobactam biosynthesis	map00261	1	ec:2.7.2.4 - kinase
9	alpha-Linolenic acid metabolism	map00592	1	ec:1.3.3.6 - oxidase
10	Lipopolysaccharide biosynthesis	map00540	1	ec:2.5.1.55 - synthase
11	Valine, leucine and isoleucine biosynthesis	map00290	2	ec:2.6.1.42 - transaminase, ec:4.2.1.9 - dehydratase
12	Styrene degradation	map00643	1	ec:3.5.1.4 - acylamidase

13	Glycosphingolipid biosynthesis - ganglio series	map00604	1	ec:3.2.1.52 - hexosaminidase
14	Valine, leucine and isoleucine degradation	map00280	1	ec:2.6.1.42 - transaminase
15	Nitrotoluene degradation	map00633	1	ec:2.3.1.5 - N-acetyltransferase
16	Amino sugar and nucleotide sugar metabolism	map00520	1	ec:3.2.1.52 - hexosaminidase
17	Biosynthesis of unsaturated fatty acids	map01040	1	ec:1.3.3.6 - oxidase
18	N-Glycan biosynthesis	map00510	1	ec:3.2.1.106 - glucosidase
19	Phenylalanine metabolism	map00360	1	ec:3.5.1.4 - acylamidase
20	Atrazine degradation	map00791	1	ec:3.5.1.5 - ec:3.5.1.5 urease
21	Sphingolipid metabolism	map00600	1	ec:3.1.6.1 - sulfatase
22	Aminoacyl-tRNA biosynthesis	map00970	1	ec:6.1.1.14 - ligase
23	Aminobenzoate degradation	map00627	1	ec:3.5.1.4 - acylamidase
24	Oxidative phosphorylation	map00190	2	ec:1.3.5.1 - dehydrogenase, ec:1.6.99.3 - dehydrogenase
25	Cysteine and methionine metabolism	map00270	3	ec:2.6.1.42 - transaminase, ec:2.7.2.4 - kinase, ec:2.3.1.30 - O-acetyltransferase
26	Folate biosynthesis	map00790	1	ec:4.1.99.22 - 3',8-cyclase
27	Glycine, serine and threonine metabolism	map00260	1	ec:2.7.2.4 - kinase
28	Fatty acid degradation	map00071	1	ec:1.3.3.6 - oxidase
29	Starch and sucrose metabolism	map00500	2	ec:3.2.1.2 - saccharogen amylase, ec:2.4.1.34 - synthase
30	Drug metabolism - other enzymes	map00983	5	ec:3.1.1.1 - ali-esterase, ec:1.17.3.2 - oxidase, ec:1.17.4.1 - reductase, ec:2.7.1.21 - inase, ec:2.3.1.5 - N-acetyltransferase
31	Alanine, aspartate and glutamate metabolism	map00250	1	ec:3.5.1.2 - glutaminase I
32	Fatty acid biosynthesis	map00061	4	ec:2.3.1.85 - synthase system, ec:6.4.1.2 - carboxylase, ec:2.3.1.180 - synthase III, ec:2.3.1.41 - synthase I
33	Glycosphingolipid biosynthesis - globo and isoglobo series	map00603	1	ec:3.2.1.52 - hexosaminidase
34	Pyrimidine metabolism	map00240	2	ec:1.17.4.1 - reductase, ec:2.7.1.21 - kinase
35	Fructose and mannose metabolism	map00051	1	ec:5.3.1.5 - isomerase
36	Aflatoxin biosynthesis	map00254	1	ec:6.4.1.2 - carboxylase
37	Arginine and proline metabolism	map00330	2	ec:3.5.1.4 - acylamidase, ec:1.2.1.71 - dehydrogenase
38	Biotin metabolism	map00780	3	ec:2.3.1.41 - synthase I, ec:2.6.1.62 - transaminase, ec:6.3.3.3 - synthase
39	Various types of N-glycan biosynthesis	map00513	1	ec:3.2.1.52 - hexosaminidase
40	Lysine biosynthesis	map00300	1	ec:2.7.2.4 - kinase
41	Sulfur metabolism	map00920	1	ec:2.3.1.30 - O-acetyltransferase
42	Pantothenate and CoA biosynthesis	map00770	3	ec:2.6.1.42 - transaminase, ec:1.1.1.169 - 2-reductase, ec:4.2.1.9 - dehydratase
43	Steroid hormone biosynthesis	map00140	1	ec:3.1.6.1 - sulfatase
44	Purine metabolism	map00230	10	ec:3.6.1.15 - phosphatase, ec:2.7.6.5 - diphosphokinase, ec:3.5.1.5 - ec:3.5.1.5 urease, ec:3.6.1.11 - metaphosphatase, ec:1.17.1.4 - dehydrogenase, ec:3.5.4.3 - deaminase, ec:1.17.3.2 - oxidase, ec:1.17.4.1 - reductase, ec:3.6.1.3 - adenylpyrophosphatase, ec:6.3.5.3 - synthase
45	Glycerolipid metabolism	map00561	1	ec:2.3.1.20 - O-acyltransferase
46	Arginine biosynthesis	map00220	4	ec:2.3.1.1 - N-acetyltransferase, ec:3.5.1.5 - ec:3.5.1.5 urease, ec:2.3.1.35 - N-acetyltransferase, ec:3.5.1.2 - glutaminase I
47	Methane metabolism	map00680	2	ec:4.1.1.31 - carboxylase, ec:6.2.1.1 - ligase
48	Biosynthesis of secondary metabolites - unclassified	map00999	1	ec:2.3.1.30 - O-acetyltransferase

49	Pentose and glucuronate interconversions	map00040	1	ec:5.3.1.5 - isomerase
50	D-Glutamine and D-glutamate metabolism	map00471	1	ec:3.5.1.2 - glutaminase I
51	Ubiquinone and other terpenoid-quinone biosynthesis	map00130	1	ec:2.1.1.201 - methylase
52	Butanoate metabolism	map00650	1	ec:1.3.5.1 - dehydrogenase
53	Glucosinolate biosynthesis	map00966	1	ec:2.6.1.42 - transaminase
54	Riboflavin metabolism	map00740	1	ec:2.5.1.78 - synthase
55	Glycerophospholipid metabolism	map00564	2	ec:3.1.1.5 - lecithinase B, ec:1.1.5.3 - dehydrogenase
56	Thiamine metabolism	map00730	1	ec:3.6.1.15 - phosphatase
57	Cutin, suberine and wax biosynthesis	map00073	1	ec:2.3.1.20 - O-acyltransferase
58	Carbon fixation pathways in prokaryotes	map00720	4	ec:1.3.5.1 - dehydrogenase, ec:4.1.1.31 - carboxylase, ec:6.4.1.2 - carboxylase, ec:6.2.1.1 - ligase
59	Glycosaminoglycan degradation	map00531	2	ec:3.1.6.14 - chondroitinsulfatase, ec:3.2.1.52 - hexosaminidase
60	Carbon fixation in photosynthetic organisms	map00710	1	ec:4.1.1.31 - carboxylase
61	Pentose phosphate pathway	map00030	2	ec:1.1.1.49 - dehydrogenase (NADP+), ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating)
62	Other glycan degradation	map00511	1	ec:3.2.1.52 - hexosaminidase
63	Peptidoglycan biosynthesis	map00550	1	ec:3.4.16.4 - D-Ala-D-Ala carboxypeptidase
64	Citrate cycle (TCA cycle)	map00020	2	ec:1.3.5.1 - dehydrogenase, ec:4.1.1.32 - carboxykinase
65	Propanoate metabolism	map00640	3	ec:6.4.1.2 - carboxylase, ec:4.1.3.30 - lyase, ec:6.2.1.1 - ligase

Table S7. 6. Enzymes involved in the biosynthesis of secondary metabolites.

	Enzyme	Pathway	Pathway ID	Enzymes in in Pathway
1	ec:1.1.1.169 - 2-reductase	Pantothenate and CoA biosynthesis	map00770	3
2	ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating)	Pentose phosphate pathway, Glutathione metabolism, Biosynthesis of antibiotics	map00030, map00480, map01130	2, 4, 13
3	ec:1.1.1.49 - dehydrogenase (NADP+)	Pentose phosphate pathway, Glutathione metabolism, Biosynthesis of antibiotics	map00030, map00480, map01130	2, 4, 13
4	ec:1.1.5.3 - dehydrogenase	Glycerophospholipid metabolism	map00564	2
5	ec:1.11.1.15 - thioredoxin peroxidase	Glutathione metabolism	map00480	4
6	ec:1.17.1.4 - dehydrogenase	Purine metabolism	map00230	10
7	ec:1.17.3.2 - oxidase	Drug metabolism - other enzymes, Purine metabolism, Caffeine metabolism	map00983, map00230, map00232	5, 10, 2
8	ec:1.17.4.1 - reductase	Glutathione metabolism, Pyrimidine metabolism, Drug metabolism - other enzymes, Purine metabolism	map00480, map00240, map00983, map00230	4, 2, 5, 10
9	ec:1.2.1.71 - dehydrogenase	Arginine and proline metabolism	map00330	2
10	ec:1.3.3.6 - oxidase	Biosynthesis of unsaturated fatty acids, Fatty acid degradation, alpha-Linolenic acid metabolism	map01040, map00071, map00592	1, 1, 1
11	ec:1.3.5.1 - dehydrogenase	Carbon fixation pathways in prokaryotes, Oxidative phosphorylation, Citrate cycle (TCA cycle), Biosynthesis of antibiotics, Butanoate metabolism	map00720, map00190, map00020, map01130, map00650	4, 2, 2, 13, 1
12	ec:1.6.99.3 - dehydrogenase	Oxidative phosphorylation	map00190	2

13	ec:2.1.1.201 - methylase	Ubiquinone and other terpenoid-quinone biosynthesis	map00130	1
14	ec:2.3.1.1 - N-acetyltransferase	Biosynthesis of antibiotics, Arginine biosynthesis	map01130, map00220	13, 4
15	ec:2.3.1.180 - synthase III	Fatty acid biosynthesis	map00061	4
16	ec:2.3.1.20 - O-acyltransferase	Cutin, suberine and wax biosynthesis, Glycerolipid metabolism	map00073, map00561	1, 1
17	ec:2.3.1.30 - O-acetyltransferase	Biosynthesis of secondary metabolites - unclassified, Sulfur metabolism, Cysteine and methionine metabolism, Biosynthesis of antibiotics	map00999, map00920, map00270, map01130	1, 1, 3, 13
18	ec:2.3.1.35 - N-acetyltransferase	Biosynthesis of antibiotics, Arginine biosynthesis	map01130, map00220	13, 4
19	ec:2.3.1.41 - synthase I	Fatty acid biosynthesis, Biotin metabolism	map00061, map00780	4, 3
20	ec:2.3.1.5 - N-acetyltransferase	Drug metabolism - other enzymes, Caffeine metabolism, Nitrotoluene degradation	map00983, map00232, map00633	5, 2, 1
21	ec:2.3.1.85 - synthase system	Fatty acid biosynthesis	map00061	4
22	ec:2.4.1.34 - synthase	Starch and sucrose metabolism	map00500	2
23	ec:2.5.1.55 - synthase	Lipopolysaccharide biosynthesis	map00540	1
24	ec:2.5.1.78 - synthase	Riboflavin metabolism	map00740	1
25	ec:2.6.1.42 - transaminase	Valine, leucine and isoleucine biosynthesis, Cysteine and methionine metabolism, Valine, leucine and isoleucine degradation, Pantothenate and CoA biosynthesis, Glucosinolate biosynthesis, Biosynthesis of antibiotics	map00290, map00270, map00280, map00770, map00966, map01130	2, 3, 1, 3, 1, 13
26	ec:2.6.1.62 - transaminase	Biotin metabolism	map00780	3
27	ec:2.7.1.21 - kinase	Pyrimidine metabolism, Drug metabolism - other enzymes	map00240, map00983	2, 5
28	ec:2.7.2.4 - kinase	Glycine, serine and threonine metabolism, Cysteine and methionine metabolism, Lysine biosynthesis, Biosynthesis of antibiotics, Monobactam biosynthesis	map00260, map00270, map00300, map01130, map00261	1, 3, 1, 13, 1
29	ec:2.7.6.5 - diphosphokinase	Purine metabolism	map00230	10
30	ec:3.1.1.1 - ali-esterase	Drug metabolism - other enzymes	map00983	5
31	ec:3.1.1.5 - lecithinase B	Glycerophospholipid metabolism	map00564	2
32	ec:3.1.6.1 - sulfatase	Steroid hormone biosynthesis, Sphingolipid metabolism	map00140, map00600	1, 1
33	ec:3.1.6.14 - chondroitinsulfatase	Glycosaminoglycan degradation	map00531	2
34	ec:3.2.1.106 - glucosidase	N-Glycan biosynthesis	map00510	1
35	ec:3.2.1.2 - saccharogen amylase	Starch and sucrose metabolism	map00500	2
36	ec:3.2.1.52 - hexosaminidase	Glycosphingolipid biosynthesis - ganglio series, Amino sugar and nucleotide sugar metabolism, Other glycan degradation, Glycosphingolipid biosynthesis - globo and isoglobo series, Glycosaminoglycan degradation, Various types of N-glycan biosynthesis	map00604, map00520, map00511, map00603, map00531, map00513	1, 1, 1, 1, 2, 1
37	ec:3.4.16.4 - D-Ala-D-Ala carboxypeptidase	Peptidoglycan biosynthesis	map00550	1
38	ec:3.5.1.2 - glutaminase I	Alanine, aspartate and glutamate metabolism, D-Glutamine and D-glutamate metabolism, Arginine biosynthesis	map00250, map00471, map00220	1, 1, 4
39	ec:3.5.1.4 - acylamidase	Tryptophan metabolism, Phenylalanine metabolism, Arginine and proline metabolism, Styrene degradation, Aminobenzoate degradation	map00380, map00360, map00330, map00643, map00627	1, 1, 2, 1, 1
40	ec:3.5.1.5 - ec:3.5.1.5 urease	Atrazine degradation, Purine metabolism, Arginine biosynthesis	map00791, map00230, map00220	1, 10, 4

41	ec:3.5.4.3 - deaminase	Purine metabolism	map00230	10
42	ec:3.6.1.11 - metaphosphatase	Purine metabolism	map00230	10
43	ec:3.6.1.15 - phosphatase	Purine metabolism, Thiamine metabolism	map00230, map00730	10, 1
44	ec:3.6.1.3 - adenylpyrophosphatase	Purine metabolism	map00230	10
45	ec:4.1.1.31 - carboxylase	Carbon fixation pathways in prokaryotes, Carbon fixation in photosynthetic organisms, Pyruvate metabolism, Methane metabolism	map00720, map00710, map00620, map00680	4, 1, 4, 2
46	ec:4.1.1.32 - carboxykinase (GTP)	Pyruvate metabolism, Citrate cycle (TCA cycle), Glycolysis / Gluconeogenesis, Biosynthesis of antibiotics	map00620, map00020, map00010, map01130	4, 2, 2, 13
47	ec:4.1.3.1 - lyase	Glyoxylate and dicarboxylate metabolism	map00630	2
48	ec:4.1.3.30 - lyase	Propanoate metabolism	map00640	3
49	ec:4.1.99.22 - 3',8-cyclase	Folate biosynthesis	map00790	1
50	ec:4.2.1.9 - dehydratase	Valine, leucine and isoleucine biosynthesis, Pantothenate and CoA biosynthesis, Biosynthesis of antibiotics	map00290, map00770, map01130	2, 3, 13
51	ec:5.3.1.5 - isomerase	Pentose and glucuronate interconversions, Fructose and mannose metabolism	map00040, map00051	1, 1
52	ec:6.1.1.14 - ligase	Aminoacyl-tRNA biosynthesis	map00970	1
53	ec:6.2.1.1 - ligase	Carbon fixation pathways in prokaryotes, Propanoate metabolism, Pyruvate metabolism, Methane metabolism, Glycolysis / Gluconeogenesis, Biosynthesis of antibiotics, Glyoxylate and dicarboxylate metabolism	map00720, map00640, map00620, map00680, map00010, map01130, map00630	4, 3, 4, 2, 2, 13, 2
54	ec:6.3.3.3 - synthase	Biotin metabolism	map00780	3
55	ec:6.3.5.3 - synthase	Purine metabolism, Biosynthesis of antibiotics	map00230, map01130	10, 13
56	ec:6.4.1.2 - carboxylase	Fatty acid biosynthesis, Carbon fixation pathways in prokaryotes, Aflatoxin biosynthesis, Propanoate metabolism, Pyruvate metabolism, Biosynthesis of antibiotics	map00061, map00720, map00254, map00640, map00620, map01130	4, 4, 1, 3, 4, 13

Figure S7. 1. Manhattan plots for the twelve biochemical traits based on the MLM model.

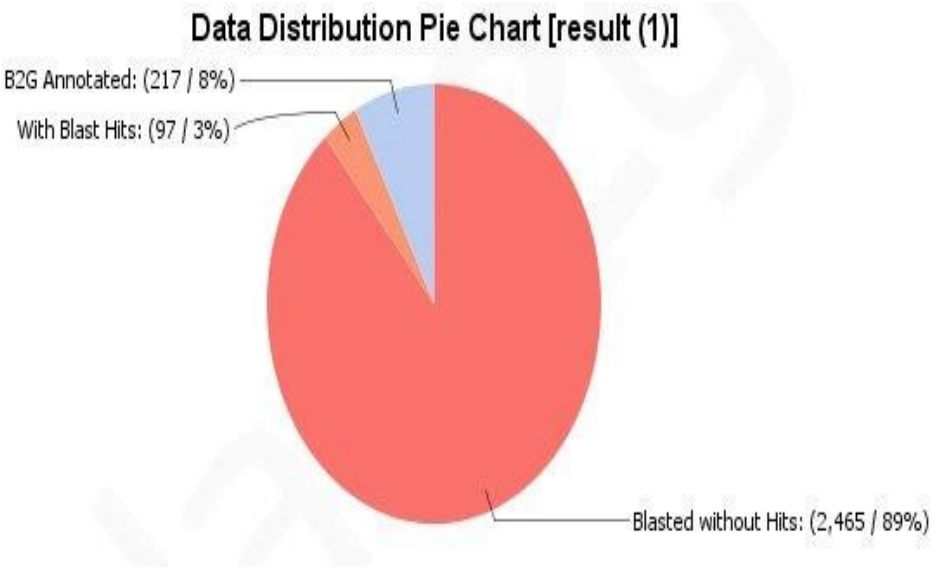

Figure S7. 2. Data distribution pie chart.

General conclusion

Tea quality is an important trait that influences the pricing of tea at the auction. However, evaluating tea quality is a major challenge facing the tea industry and tea breeding programmes. Miniature manufacture of tea is typically the first step prior to evaluating black tea quality in a breeding programme. Processing of tea is complicated by variabilities such as genotype, post-harvest handling and environmental conditions, that are challenging to control. Our results showed that it is possible to select high quality teas in a breeding programme using only fresh leaf biochemical data; catechins, caffeine and theanine. Sensory evaluation using professional tea tasters is traditionally used to determine the quality and pricing of made tea. However, it is highly subjective. Therefore, analytical methods and sensory bionic techniques could be integrated in routine tea quality determination commercially and in tea breeding programmes, because they are objective, and the results positively correlate with those of professional tea tasters.

Conventional tea breeding using field selection and hybridization is the main method used to breed tea. However, it has many weaknesses which marker assisted selection could help to overcome. The draft genomes of *Camellia sinensis* var. *sinensis* and *Camellia sinensis* var. *assamica* have recently been sequenced and published and could provide new insights to be explored, to further improve tea quality. Additionally, the development of high-throughput DNA sequencing technologies such as SNP genotyping and genotyping-by-sequencing (GBS), allows rapid genotyping of many genotypes at a lower cost.

We identified several markers and candidate genes associated with tea quality using GBS markers and this provide valuable resources for future studies of tea quality thus creating opportunities for accelerating molecular tea breeding. Our study provides the first application of GS using GBS for predicting high quality teas in tea breeding programmes. A GS framework that could be implemented in tea has been proposed. While model choice and a suitable genotyping approach are key factors for implementing GS, the genetic architecture of the trait, heritability, marker density, LD between the QTL and the markers, size of the training population and the relationship between the training and validation populations also play a

significant role in making decisions. We hope that implementing GS in breeding for high quality tea varieties will result in higher prediction accuracy and rapid gains from selection.

Recommendations for future study

Our results demonstrated that biochemical compounds can be used to select high quality teas. This will improve the accuracy and hence reduce the subjectivity associated with organoleptic evaluation. Analysis of the major catechins revealed that EGCG and EGC correlated with taste, while ECG correlated with mouthfeel, and could consequently be used to select high quality teas. Taste and mouthfeel are important sensory traits that professional tasters use to grade tea at the auction. However, only eight contrasting varieties were used in this study. Therefore, we recommend validation of these results using many different genetically diverse genotypes. This will inform the future of tea quality evaluation.

We identified several markers and potential candidate genes associated with tea quality. However, these results should be taken with caution because our ability is limited by the resolution of the GBS markers and the current draft tea genome assembly. Our study used GBS to provide the markers for GS and GWAS analyses at a reasonable cost. However, the use of proper missing marker imputation algorithms is necessary to overcome the generally observed problem of missing marker data with GBS. Additionally, all the 2779 markers used in this study should be validated and the identified significant SNPs should be fine mapped further to assess their potential involvement in tea quality.

This study was initiated before the release of the draft tea genomes. However, as the assemblies of the two draft genomes are not ordered along the chromosomes, they provide little benefit over de novo SNP marker discovery used in this study. Additionally, the two draft genomes have many sequence-repeat regions and scaffolds, making it difficult to identify candidate genes of interest by location. Therefore, further re-sequencing is necessary to improve the genome assembly and content of the current tea genomes.